Robust Approach to Risk Management and Statistical Analysis

WONG, Man Hong

A Thesis Submitted in Partial Fulfilment

of the Requirements for the Degree of

Doctor of Philosophy

in

Systems Engineering and Engineering Management

The Chinese University of Hong Kong

June 2012

Thesis/Assessment Committee

Professor LI Duan (Chair)

Professor ZHANG Shuzhong (Thesis Supervisor)

Professor SO Man Cho (Thesis Supervisor)

Professor CHEN Nan (Committee Member)

Doctor DANG Chuangyin (External Examiner)

Abstract In this thesis we study some structural results in polynomial optimization, with an emphasis paid to the applications from risk management problems and estimations in statistical analysis. The key underlying method being studied is related to the so-called S-lemma in control theory and robust optimization. The original S-lemma was developed by Yakubovich, which states an equivalent condition for a quadratic polynomial to be non-negative over the non-negative domain of other quadratic polynomial(s). In this thesis, we extend the S-Lemma to univariate polynomials of any degree. Since robust optimization has a strong connection to the S-Lemma, our results lead to many applications in risk management and statistical analysis, including estimating certain nonlinear risk measures under moment bound constraints, and an SDP formulation for simultaneous confidence bands. Numerical experiments are conducted and presented to illustrate the effectiveness of the methods.

摘要　**本**博士論文著重研究關於多項式優化的理論，並討論其在風險管理及統計分析中的應用。我們主要研究的對象乃為在控制理論和穩健優化中常見的所謂Ｓ引理。 原始的Ｓ引理最早由Ｙａｋｕｂｏｖｉｃｈ所引入。它給出一個二次多項式在另一個二次多項式的非負域上為非負的等價條件。在本論文中，我們把Ｓ引理推廣到一元高次多項式。 由於Ｓ引理與穩健優化密切相關，所以我們的結果可廣泛應用於風險管理及統計分析，包括估算在高階矩約束下的非線性風險量度問題，以及利用半正定規劃來計算同時置信區域帶等重要課題。 同時，在相關章節的末段，我們以數值實驗結果來引證有關的新理論的有效性和應用前景。

Acknowledgement

I would like to thank my supervisors, Professor Shuzhong ZHANG and Professor Anthony SO, for their guidance and kind support for my research during my PhD study. Throughout the years I have learnt from them not only the technical knowledge but also the philosophy of research and life. Special thanks are due to Professor Simai HE, who has helped and advised me in a substantial way on my research leading to the results contained in this thesis.

This work is dedicated to my family, especially my wife, for their love and support. Without them, this thesis would not have been possible.

Contents

List of Tables

Chapter 1

Introduction

Robust optimization has found many applications in risk management and engineering because of its wide spread modeling power in solving decision problems under uncertain parameters. In many cases, the so-called S-Lemma plays a significant role in the modeling process since it is capable of formulating some specific types of semi-infinite optimization models by a finite number of linear matrix inequalities (LMIs). As a consequence, the S-Lemma has attracted much research interest in itself and a large amount of follow-up work can be found in the literature. A significant portion of this thesis will be devoted to discussing the S-Lemma under the context of univariate polynomials, which is a novel extension of the classical S-Lemma. Apart from the theory, several robustness issues in statistical analysis and risk management are discussed.

Throughout the thesis except the last chapter, our discussion is focused on the class of univariate polynomials (unless otherwise specified). Chapters 2 to 4 contain the theoretical developments, while Chapters 5 to 7 are related to the applications in practice, which arise in some specific robust optimization formulations. More specifically, Chapter 2 discusses the literature on the S-Lemma and its background, followed by some motivating examples. Chapter 3 is concerned about the application of the new variant of the S-Lemma in practice; in fact, we shall present a strengthened robust condition, which leads to reasonable and tractable robust optimization formulations.

In Chapter 4, the new variant of the S-Lemma for the univariate polynomials will be presented. That chapter mainly contains a series of lemmas and theorems. In Chapter 5, with the newly established S-Lemma, we tackle a new class of problems arising from applications in statistics: confidence bands for polynomial regression. While there are methodologies for the computation of the confidence bands, we present a new approach with our new S-Lemma through semidefinite programming (SDP). Numerical experiments are conducted to support our formulations towards the end of the chapter.

As another relevant application, moment bound of nonlinear risk is discussed in Chapter 6. Extreme events occur rarely, but these are often the circumstance where an insurance coverage is needed. Given the first, say, n moments of the risk(s) of the events, one is able to compute or approximate the tight bounds for a risk measures in the form of $\mathbb{E}(\psi(x))$ through SDP, via distributional robust optimization formulations. Existing results in the literature have already demonstrated the power of this technique when $\psi(x)$ is linear or piecewise linear. We extend the method to the case where $\psi(x)$ is a polynomial or fractional of polynomials.

In Chapter 7, we introduce the issue of calculating the moment bound for worst-case joint probability with two random variables. Consider a random vector, and assume that a set of its moments information is known. Among all possible distributions obeying the given moments constraints, the envelope of the probability distribution functions is introduced in this chapter as distributional robust probability function. We show that such a function is computable in the bi-variate case under some conditions. The same chapter also includes a discussion on the connections to the existing results in the literature, as well as the applications in risk management.

□ End of chapter.

Chapter 2

Meeting the S-Lemma

The fundamental principle of the S-Lemma (or the S-procedure) is to support the equivalence of the following two statements.

Statement 1.

$$f(x) \geq 0 \quad \forall g_i(x) \geq 0, i = 1, \cdots, k. \tag{2.1}$$

Statement 2. *There exists nonnegative constants* $\lambda_1, \cdots, \lambda_k$*, not all zero, such that*

$$f(x) - \sum_{i=0}^{k} \lambda_i g_i(x) \geq 0. \tag{2.2}$$

It is easy to verify that Statement 1 is a consequence of Statement 2. If the reverse holds, then it is called *lossless*, in which case we say that the S-Lemma is applicable. An informative review of the S-Lemma as well as its application in various fields of mathematics (functional analysis, rank-constrained optimization and generalized convexities) can be found in Polik and Terlaky [57] and the references therein. When $k = 1$ and the discussion is limited to quadratic functions, Yakubovich [80, 81] first proved the equivalence with a regularity condition:

Theorem 1. *Let $f, g : \mathbf{R}^n \mapsto \mathbf{R}$ be quadratic functions and suppose that there is an $\bar{x} \in \mathbf{R}^n$ such that $g(\bar{x}) < 0$. Then the following two statements are equivalent.*

1. *There is no $x \in \mathbf{R}^n$ such that*

$$\begin{cases} f(x) < 0 \\ g(x) \geq 0 \end{cases}$$

2. *There is a nonnegative number $y \geq 0$ such that*

$$f(x) + yg(x) \geq 0 \forall\, x \in \mathbf{R}^n$$

The proof is based on a convexity result of Dines [20]:

Theorem 2. *(Dines [20]; see also Polik and Terlaky [57]) If $f, g : \mathbf{R}^n \mapsto \mathbf{R}$ are homogeneous quadratic functions, then the set $\{(f(x), g(x) : x \in \mathbf{R}\} \subset \mathbf{R}^2$ is convex.*

There are alternative proofs for the S-Lemma found in the literature. One is in Ben-Tal and Nemirovski [4] which relies on a technical lemma of Sturm and Zhang [70] regarding a constructive procedure for a specific rank-1 decomposition of a semidefinite matrix. Another version is based on Lemma 2.3 in Yuan [82]:

Theorem 3. *(Yuan [82]; see also Ai et al. [1])Let A_1 and A_2 be positive semidefinite matrices in $\mathbf{R}^{n\times n}$. Then the following are equivalent:*

1. $\max\{x^T A_1 x, x^T A_2 x\} \geq 0$ *for all* $x \in \mathbf{R}^n$ *(resp.* > 0 *for all* $x \neq 0$*).*

2. *There exist* $\mu_1 \geq 0, \mu_2 \geq 0, \mu_1 + \mu_2 = 1$ *such that* $\mu_1 A_1 + \mu_2 A_2$ *is positive semidefinite matrices in* $\mathbf{R}^{n\times n}$ *(resp. positive definite).*

Efforts have been made to explore the conditions under which Statement 1 and Statement 2 are equivalent for a larger k, but there are rarely any successful cases. Polyak [59] discusses the case $k = 2$ with a further assumption. Recently, Hu and Huang [31] provide the same result based on a different assumption and generalized S-Lemma to even order tenors. Meanwhile, Tuy and Tuan [74] apply the topological minimax theorem to establish a generalized S-Lemma (k can be any integer) that x is restricted to an affine manifold. The results shed some light on the relationship between the S-Lemma and the saddle point theorem of the minimax formulation. On the other hand, although the

S-Lemma has been limited to quadratic functions with a single quadratic inequality $g(x)$, its footprints can be found in quite a variety of fields. For instance, stability analysis of a dynamic system using Lyapunov functions (see Jonsson [34]), estimating sum of ellipsoids in error estimating (see Polyak [59]), and applications in robust optimizations (see Chen et al. [10]).

Given the broad interest and wide applicability of the S-Lemma, it is natural to extend the equivalence between Statements 1 and 2 in various circumstances. One such natural extension is to consider univariate polynomials. In the following, some motivating applications are presented to illustrate the point.

Example 1. ***(Cash Flow Management)***
In asset-liability management, we are often required to allocate cash flows in different future periods in order to meet a set of obligations. Suppose we have to decide the cash flow for a series of obligation O_i at the end of period $i = 0, \cdots, n$. Meanwhile, a series of cash flow, $b_1, \cdots, b_n$ has already been arranged. In case some of cash flows are positive, they can be used to fund the obligations. In order to justify the whole portfolio economically (i.e. a nonnegative future value) under different interest rate environments, we can formulate an optimization problem to find

the minimal amount of cash required for each period:

$$\begin{aligned} &\min \sum_{i=0}^{n} a_i \\ &s.t. \sum_{i=0}^{n} (a_i + b_i - O_i)(1+x)^{n-i} \geq 0 \quad \forall x \in [r_1, r_2] \cup [r_3, r_4], \end{aligned}$$

where a_i's are the decision variables, and $[r_1, r_2]$ and $[r_3, r_4]$ (with $r_1 < r_2 < r_3 < r_4$) reflect different economic environments, say boom and bust. Recall that $x \in [r_1, r_2] \cup [r_3, r_4]$ can be represented by a polynomial; thus the constraint is an example of Statement 1.

Example 2. ***("Best-fit" Polynomial)***
How functions can be best approximated by polynomials is of central importance in approximation theory. This is achieved by minimizing the quantified error(s) between the given functions and the approximating polynomial. This idea can be illustrated by polynomial regressions.

Given a set of sample points, we can regress it with a polynomial $\hat{f}_i(x)$. If trials are repeated M times, we obtain $N(\leq M)$ possibly different justified $\hat{f}_i(x)$'s. Applying the concept in approximation theory, we can formulate an optimization problem for a "best-fit" polynomial based on the $\hat{f}_i(x)$'s: $\min \sum_{i=1}^{N} |f(x) - \hat{f}_i(x)|$. In general, the approximation is measured on a particular range $\mathcal{I}$ (that can be an interval or a union of such). Then *the formulation is*

$$\begin{aligned}
&\min \sum_{i=1}^{N} t_i \\
&s.t. \quad -t_i \le f(x) - \hat{f}_i(x) \le t_i \quad \forall x \in \mathcal{I}, i = 1, \cdots, N
\end{aligned}$$

It can be easily seen that the constraints are examples of Statement 1

Example 3. ***(Moment Bounds)***

When we estimate the extremal (say maximum) expectation of a certainty quantity $\psi(x)$ based on the moments information of its underlying randomness $x \in \Omega$, it is called a moment bound problem. The estimation is often through its dual formulation. Assuming $\Omega \subseteq \mathbf{R}$, the problem can be formulated as follows:

$$\begin{aligned}
\sup \mathbb{E}[\psi(x)] = &\inf_{z_0, \cdots, z_n} \sum_{i=0}^{n} m_i z_i \\
&s.t. \quad \sum_{i=0}^{n} z_i x^i \ge \psi(x) \quad \forall x \in \Omega,
\end{aligned}$$

where m_i are the i-th moment of x. Given more information about the range of x, the "for-all" ($\forall$) condition can be represented by some nonnegative polynomial. Then the constraints is an instance of Statement 1, provided that $\psi(x)$ is also a (piecewise) polynomial.

However, a straight-forward analogy of S-Lemma in univariate polynomials does not apply. In other words, there cannot be such nonnegative constant y in Statement 2 in general, even in the case of univariate polynomials. Counter examples can be easily constructed with odd degree polynomials, as well as when both $f(x)$ and $g(x)$ are even degree polynomials. In all cases, there is no guarantee for the existence of a nonnegative constant y. The following examples can be used to illustrate the point.

Example 4. *Consider the following pairs of polynomials*

1. $f(x) = (x+1)^2(x-2)(x-4)$ *and* $g(x) = x(x+2)(x-6)$
2. $f(x) = (x+4)(x+3)(x+1)^2(x-7)$ *and* $g(x) = x(x+2)(x-1)(x-4)(x-10)$
3. $f(x) = (x+2)(x+3)(x-8)(x-10)$ *and* $g(x) = x(x+5)(x-1)(x-5)$

We can verify the existence of y *by solving an optimization problem:* $\max_{f(x)-yg(x)\geq 0} y$*. The optimal value however is* $-\infty$*. This means that the original S-Lemma fails if* $f(x)$ *and* $g(x)$ *are not quadratic functions.*

Since requiring y to be a constant is too strong for the statements to be equivalent, even for univariate polynomials, we introduce a more relaxed statement to replace Statement 2:

Statement 3. *There exist positive function $\lambda_0(x)$ and nonnegative functions $\lambda_i(x)$, $i = 1, \cdots, k$, not all identically zero, such that*

$$\lambda_0(x)f(x) - \sum_{i=1}^{k} \lambda_i(x)g_i(x) \geq 0 \quad \forall\, x \in \mathbf{R} \tag{2.3}$$

Note that when $\lambda_0(x) \equiv 1$ and all $\lambda_i(x)$ are constants, Statement 3 reduces to Statement 2. From now on, we discuss only univariate polynomials. In the next chapter, we will first discuss a stronger condition guaranteeing the implications of the statements. Then in Chapter 4, we are going to establish the general equivalence between Statements 1 and 3, with a regularity condition.

□ **End of chapter.**

Chapter 3

A strongly robust formulation

Before we prove the existence of $h_1(x)$ and $h_2(x)$ in Statement 3 to establish the S-Lemma for univariate polynomials in the next chapter, let us consider a stronger robust formulation.

3.1 A more practical extension for robust optimization

3.1.1 Motivation from modeling aspect

In robust optimization,

$$\begin{aligned} \min \quad & c^T\eta \\ s.t. \quad & f_\eta(x) \geq 0 \; \forall \; g_i(x) \geq 0 \quad i = 1, \cdots, k \end{aligned} \tag{3.1}$$

is regarded as the robustness constraint, in the sense that $f_\eta(x)$, whose coefficients affinely depend on the decision variable η, must satisfy all the required condition (i.e. $\forall\ g_i(x) \geq 0,\ i = 1, \cdots, k$). This constraint is not tractable in general. When $k = 1$ and both $f_\eta(x)$ and $g_1(x)$ are quadratic function (and x could be in $\mathbf{R}^n$), the original S-Lemma guarantees the tractability by formulating an equivalent LMI. For $k > 1$, the S-Lemma fails to hold and we only have an SDP relaxation as an approximation (see Nemirovski [54]).

On the other hand, when $k = 1$ and both $f_\eta(x)$ and $g_1(x)$ are univariate polynomials, as is in this thesis's setting, although we confirm (in the next chapter) the existence of $h_1(x) > 0$ and $h_2(x) \geq 0$ such that $h_1(x)f_\eta(x) - h_2(x)g_1(x) \geq 0\ \forall\ x \in \mathbf{R}$, the tractability is lost. This is because we have introduced new variables in the coefficients of $h_1(x)$ so that some variables (together with η) are no longer affine linear with respect to the coefficients of the term $h_1(x)f(x)$, thus such a condition is insufficient to be formulated as an LMI. In a simpler setting, by Theorem 7, if we restrict that $\deg g_1(x) \leq 2$ (so that $\deg h_1(x) = 0$), then the tractability is resumed, since a quadratic or linear function $g_1(x)$ being nonnegative is equivalent to trapping x in a bounded or semi-infinite interval, or being free over the real line. This reduces to the discussion in Nesterov [56] about the nonnegativity of polynomial over such intervals. Nonetheless, the limitation for a general degree of $g_1(x)$ remains.

In view of the difficulties in tackling (3.1) efficiently in the polynomial case, in this chapter we will replace it by a more robust (conservative) condition for polynomials of an even degree, so that $h_1(x)$ and $h_2(x)$ can be constants. Not only will this approach alleviate the limitation mentioned above, but also it is extendable to a more general case where two or more $g_i(x)$'s are present.

3.1.2 Discussion of a more robust condition

Let $\vec{x} := (1, x, \cdots, x^n)^T$ for any $x \in \mathbf{R}$. Rewrite the polynomials $f_\eta(x) = B_\eta \bullet \vec{x}\vec{x}^T$ and $g_i(x) = A_i \bullet \vec{x}\vec{x}^T$, where A_i ($i = 1, \cdots, k$) and B_η are some (symmetric) matrices containing the coefficients of $g_i(x)$ and $f_\eta(x)$ respectively, and $\bullet$ is the matrix inner product. Let $\mathcal{X} := \{\vec{x}\vec{x}^T | x \in \mathbf{R}\}$. Note that $\mathcal{X}$ is not convex. Recall on the other hand that Yakubovich [80, 81] proved the original S-Lemma using Dines's result [20], in which Dines proved that the joint numerical range of two homogeneous quadratic function is a convex set. In the same spirit, we consider the conic hull of the joint numerical range $(B_\eta \bullet X, A_1 \bullet X, \cdots, A_k \bullet X)$ and we have

Lemma 1.

$$\begin{aligned}\mathcal{W} :=& \textit{cone conv}\{(B_\eta \bullet X, A_1 \bullet X, \cdots, A_k \bullet X) | X \in \mathcal{X}\} \\ =& \{(B_\eta \bullet X, A_1 \bullet X, \cdots, A_k \bullet X) | X \in \textit{cone conv } \mathcal{X}\}\end{aligned}$$

Proof. Since the inner product is an affine mapping, the equivalence can be verified by definition. □

With this convex cone, we are in a position to consider

$$\begin{aligned} \min \quad & c^T \eta \\ s.t. \quad & \mathcal{W} \cap \mathcal{C} = \phi \end{aligned} \tag{3.2}$$

where $\mathcal{C} := \{(u, v_1, \cdots, v_k) \in \mathbf{R}^{k+1} | u < 0, v_i > 0, i = 1, \cdots, k\}$. Essentially, we mean to replace (3.1) by (3.2), because the latter is tractable and yet more robust than the former. These will be shown in the following two theorems.

Theorem 4. *Let $\mathcal{H}^{n+1}$ be the set of Hankel matrices in $\mathbf{R}^{(n+1)\times(n+1)}$ and $\mathcal{S}_+^{n+1} \subset \mathbf{R}^{(n+1)\times(n+1)}$ be the cone of positive semidefinite matrices. Then cone conv $\mathcal{X} = \mathcal{S}_+^{n+1} \cap \mathcal{H}^{n+1}$. Hence (3.2) can be checked by the SDP:*

$$\min \quad B_\eta \bullet X \tag{3.3}$$

$$s.t. \quad A_i \bullet X \geq 0 \quad i = 1, \cdots, k \tag{3.4}$$

$$X \in \mathcal{S}_+^{n+1} \cap \mathcal{H}^{n+1} \tag{3.5}$$

$$Tr(X) = 1 \tag{3.6}$$

Proof. It is easy to see that verifying (3.2) is equivalent to checking whether

$$\min_{\substack{A_i \bullet X \geq 0 \ i=1,\cdots,k \\ X \in cone\ conv\ \mathcal{X}}} B_\eta \bullet X \tag{3.7}$$

has a nonnegative value. To see how this can be checked efficiently, the dual of *cone conv* $\mathcal{X}$ is taken into account.

Let $Y \in (cone\ conv\ \mathcal{X})^*$. By definition,

$$\begin{aligned}
& Y \bullet X \geq 0 \ \forall X \in cone\ conv\ \mathcal{X} \\
\iff & Y \bullet t \sum_\alpha \lambda_\alpha X_\alpha \geq 0 \text{ for any convex combination } \lambda_\alpha, \\
& \text{any } t > 0 \text{ and any } X_\alpha \in \mathcal{X} \\
\iff & Y \bullet X_\alpha \geq 0 \ \forall X_\alpha \in \mathcal{X} \\
\iff & Y \bullet \vec{x}\vec{x}^T \geq 0 \ \forall x \in \mathbf{R}.
\end{aligned}$$

The last line is a univariate polynomial and one can verify that $Y \in \mathcal{S}_+^{n+1} + (\mathcal{H}^{n+1})^\perp$. Hence *cone conv* $\mathcal{X} \subseteq \mathcal{S}_+^{n+1} \cap \mathcal{H}^{n+1}$. To show their equivalence, it remains to prove that *cone conv* $\mathcal{X}$ is closed. By Lemma 1 of Sturm and Zhang [70], *cl cone conv* $\mathcal{X} =$ *cone*$\{\vec{x}\vec{x}^T | \vec{x} \in cl(\mathcal{D})\}$, where $\mathcal{D} := \{(1, x, \cdots, x^n)^T | x \in \mathbf{R}\} \subset \mathbf{R}^{n+1}$ is closed since the function $x^j : \mathbf{R} \mapsto \mathbf{R}$ is continuous for any integer j. Hence *cone conv* $\mathcal{X} = \mathcal{S}_+^{n+1} \cap \mathcal{H}^{n+1}$ and we conclude that assumption (3.2) can be checked by solving (3.3)-(3.6). □

In the next theorem, we show that (3.2) has an LMI representation which is analogous to the original S-Lemma's.

Theorem 5. *(3.2) implies the following statement:*

$$\exists y_i \geq 0 \ s.t. \ f_\eta(x) - \sum_{i=1}^{k} y_i g_i(x) \geq 0 \quad \forall\, x \in \mathbf{R}. \tag{3.8}$$

Proof. Given that $\mathcal{W} \cap \mathcal{C} = \phi$, and $\mathcal{W}$ is convex by Lemma 1, there exists $\lambda_0, \lambda_1, \cdots, \lambda_k$ (not all are zero) such that

$$\lambda_0 u + \lambda_1 v_1 + \cdots + \lambda_k v_k \leq 0 \quad \forall (u, v_1, \cdots, v_k) \in \mathcal{C} \tag{3.9}$$

$$\lambda_0 B_\eta \bullet X + \lambda_1 A_1 \bullet X + \cdots + \lambda_k A_k \bullet X \geq 0 \quad \forall x \in \mathbf{R}. \tag{3.10}$$

(3.9) implies that $\lambda_0 \geq 0, \lambda_i \leq 0$ for $i = 1, \cdots, k$. We further claim that $\lambda_0 \neq 0$. Otherwise, there exists at least one $\lambda_i \neq 0$, for some $1 \leq i \leq k,$. Then (3.10) may not hold for some x, which is a contradiction. Hence, we choose $y_i = -\frac{\lambda_i}{\lambda_0}$ and this completes the proof. □

It is worth recalling that (3.8) trivially implies (3.1). Hence, guaranteed by Theorem 5, we conclude that (3.2) is a more conservative condition than (3.1).

□ **End of chapter.**

Chapter 4

Theoretical developments

In this chapter, several ordering relations are defined in order to establish the equivalence between Statements 1 and 3 for $k = 1$.

4.1 Definition of several order relations

Let $f(x)$ and $g(x)$ be two polynomials.

Definition 1. *Three ordering relations are defined as follows:*

(a) $f|_{g\geq 0} \geq 0$ *signifies the implication:* $g(x) \geq 0 \Rightarrow f(x) \geq 0$*;*

(b) $f|_{g>0} > 0$ *signifies the implication:* $g(x) > 0 \Rightarrow f(x) > 0$*;*

(c) $f|_{g\geq 0} > 0$ *signifies the implication:* $g(x) \geq 0 \Rightarrow f(x) > 0$*.*

When all the above orders hold, $f|_{g>^*0} >^* 0$ *is used to collectively represent them.*

Definition 2. *The ordering relations $\succeq$ and $\succ$ are defined as follows:*

1. *$f(x) \succeq g(x)$ signifies the fact that*

$$\exists\ h_1(x) \geq 0, h_2(x) \geq 0 \text{ and } h_3(x) \geq 0 \text{ such that}$$
$$h_1(x)f(x) = h_2(x)g(x) + h_3(x) \quad \forall x \in \mathbf{R}$$

2. *$f(x) \succ g(x)$ signifies the fact that*

$$\exists h_1(x) \geq 0, h_2(x) \geq 0 \text{ and } h_3(x) > 0 \text{ such that}$$
$$h_1(x)f(x) = h_2(x)g(x) + h_3(x) \quad \forall x \in \mathbf{R}$$

4.2 S-Lemma with a single condition $g(x) \geq 0$

The proof of our S-Lemma relies on several technical lemmas below:

Lemma 2. ***(Decomposition Lemma)*** $f|_{g>^*0} >^* 0 \Longrightarrow \exists f_i(x)$ *with* $\deg f(x) \leq 2$*, $i = 1, \cdots, I$, such that*

(a) $f(x) = f_0(x)\prod_{i=1}^{I} f_i(x)$ *and*

(b) $f_i|_{g>^*0} >^* 0$ *for all* $i = 0, 1, \ldots, I$ *and*

(c) $f_0(x) > 0$ *for all* $x \in \mathbb{R}$.

Proof. We prove this lemma by inductively reducing the number of roots of f, based on the fact that for any two polynomial functions $\theta_1(x)|\theta_2(x)$, $\theta_1|_{g>^*0} >^* 0$ and $\theta_2|_{g>^*0} >^* 0$, we always have $\frac{\theta_2(x)}{\theta_1(x)}|_{g>^*0} >^* 0$. If we can find $\hat{f}(x)|f(x)$ with $\hat{f}|_{g>^*0} >^* 0$, $1 \leq \deg \hat{f}(x) \leq 2$, then there is a decomposition for f if and only if there is a decomposition for $f/\hat{f}$.

If $f(x)$ has no real root, then either $f > 0$ or $f < 0$. In the first case, the decomposition is trivial. In the second case, g has to be negative, therefore $f = f_0 f_1$ with $f_0 = -f$ and $f_1 = -1$ is a decomposition.

If x_0 is a real root of $f(x)$ with even multiplicities, then $(x - x_0)^2 > 0\ \forall x \neq x_0$ and $g(x_0) < 0$, therefore $(x - x_0)^2|_{g>^*0} >^* 0$ holds trivially. Applying the inductive process on $f(x)/(x-x_0)^2$, we can further decompose it until there is no multiple root.

Assume x_0 to be a real root of $f(x)$, if it is a multiple root, then $(x - x_0)^2 > 0\ \forall x \neq x_0$ and $g(x_0) < 0$, therefore $(x - x_0)^2|_{g>^*0} >^* 0$ holds trivially. Note that $\frac{f(x)}{(x-x_0)^2}|_{g>^*0} >^* 0$, we can further decompose the polynomial function $f(x)/(x - x_0)^2$ by induction. If x_0 is a single root, assume $\{x|g(x) \geq 0\} = \bigcup_{r=1}^{R}[a_r, b_r]$, where $a_r \leq b_r < a_{r+1}$ (a_1 can be $-\infty$ and b_R can be $+\infty$). Define $b_0 := -\infty$ if a_1 is bounded, and $a_{R+1} := +\infty$ if b_R is bounded. If x_0 falls in the same interval $[b_r, a_{r+1}]$ $(0 \leq r \leq R)$

with another root x_1 of $f(x)$, then $[(x-x_0)(x-x_1)]|_{g>^*0} >^* 0$ and $(x-x_0)(x-x_1)|f(x)$. Notice that there has to be an even number of roots for any bounded intervals. If there is no other root of $f(x)$ in the same interval, the interval has to be $[-\infty, a_1]$ or $[b_R, +\infty]$. For the first case, noticing that $x_0 < a_1$ and $g(x) < 0$ for all $x < a_1$, we have $(x - x_0)|_{g>^*0} >^* 0$ and $x - x_0|f(x)$. For the second case, similarly we have $(x_0 - x)|_{g>^*0} >^* 0$ and $(x_0 - x)|f(x)$. □

The contra-positivity of Definition 1 is presented as a lemma here without proof.

Lemma 3.

$$f|_{g\geq 0} > 0 \Longleftrightarrow -g|_{-f\geq 0} > 0.$$

and

$$f|_{g\geq 0} \geq 0 \Longleftrightarrow -g|_{-f>0} > 0$$

Similarly, we also have

Lemma 4. *(c.f. Lemma 3)*

$$f(x) \succeq g(x) \Longleftrightarrow -g(x) \succeq -f(x)$$

and

$$f(x) \succ g(x) \Longleftrightarrow -g(x) \succ -f(x)$$

Lemma 5. *Let* $\deg f(x) \leq 2$ *or* $f > 0$ *and* $\deg g(x) \leq 2$ *or* $-g > 0$. *Then*

(I) $f|_{g\geq 0} > 0 \Longrightarrow f(x) \succ g(x)$.

(II) $f|_{g\geq 0} \geq 0 \Longrightarrow f(x) \succeq g(x)$.

(III) $f|_{g>0} > 0 \Longrightarrow f(x) \succ g(x)$.

Furthermore, the functions h_1 *and* h_2 *corresponding to the relationship* $f(x) \succ g(x)$ *and* $f(x) \succeq g(x)$ *can be taken constants.*

Proof. Owing to the similarities in the arguments, we will only prove the lemma for (I).

If $f > 0$, then setting $h_1 = 1$, $h_2 = 0$ and $h_3 = f$ we have $h_1 f = h_2 g + h_3$. If $-g > 0$, then $h_1 = 0$, $h_2 = 1$ and $h_3 = -g$ makes $h_1 f = h_2 g + h_3$.

When $\deg f(x) = 0$ or $\deg g(x) = 0$, it is trivial.

We will explore the combinations where $\deg f(x) = 1, 2$ with $\deg g(x) = 1, 2$.

(i) $\deg f(x) = 1$ and $\deg g(x) = 1$. Let $f(x) = a_1 x + b_1$ and $g(x) = a_2 x + b_2$ with $a_1, a_2 \neq 0$. Then $f(x) = \frac{a_1}{a_2} g(x) + b_1 - \frac{a_1 b_2}{a_2}$. Take $h_1(x) = 1, h_2(x) = \frac{a_1}{a_2}$ and $h_3(x) = b_1 - \frac{a_1 b_2}{a_2}$ $|_{g\geq 0} > 0$ implies $a_1/a_2 > 0$ and $f(-b_2/a_2) > 0$. Therefore $h_2(x) \geq 0$ and $h_3(x) > 0$ for all $x \in \mathbb{R}$.

(ii) $\deg f(x) = 2$ and $\deg g(x) = 1$.

Let $f(x) = a_1x^2 + b_1x + c_1$ and $g(x) = a_2x + b_2$ with $a_1, a_2 \neq 0$. Note that $g(x) > 0$ either when x approaches $+\infty$ or $-\infty$, therefore $f(x) > 0$ when x approaches $+\infty$ or $-\infty$, which means $a_1 > 0$. Because of the affine mapping $x \to -x$, we only need to consider the case $a_2 > 0$. Let $x_0 = -\frac{b_2}{a_2}$ and $x_1 = -\frac{b_1}{2a_1}$. Then $f(x_1) = c_1 - \frac{b_1^2}{4a_1}$ and $f(x) = a_1(x - x_1)^2 + f(x_1)$. If $x_1 \geq x_0$, then we have $g(x_1) \geq 0$ so $f(x_1) > 0$. Therefore $h_1 = 1$, $h_2 = 0$ and $h_3 = f > 0$. If $x_1 < x_0$, then $f(x_0) > 0$ and $f'(x_0) > 0$. Take $h_1 = 1$, $h_2 = f'(x_0)/a_2$, and $h_3(x) = f(x) - f'(x_0)(x - x_0)$. Obviously $h_1, h_2 > 0$. For $h_3(x)$, since $f(x) \geq f'(x_0)(x - x_0) + f(x_0)$, we have $h_3(x) \geq f(x_0) > 0$ for all $x \in \mathbb{R}$. Also, it is easy to see that $h_3(x)$ is a strongly convex function because $f(x)$ is strongly convex.

(iii) $\deg f(x) = 1$ and $\deg g(x) = 2$. By Lemma 3 and Lemma 4, it follows directly from (ii).

(iv) $\deg f(x) = 2$ and $\deg g(x) = 2$. We split further our discussion into three subcases: (iv.a) $f(x)$ is convex and $g(x)$ is concave; (iv.b) both $f(x)$ and $g(x)$ are convex (when g is convex, f has to be convex); (iv.c) both $f(x)$ and $g(x)$ are concave. By Lemma 3 and Lemma 4, (iv.c) is equivalent to (iv.b). Therefore we will only discuss the first two cases.

(a) The two closed interval (maybe empty) $I_f = \{x : f(x) \leq 0\}$ and $I_g = \{x : g(x) \geq 0\}$ has to be disjoint. Therefore there exists a x_0 such that x_0 separates the two intervals. Without losing generality, we assume I_f is on the left side and I_g is on the right hand side. Then $f|_{x-x_0 \geq 0} > 0$ and $(x-x_0)|_{g\geq 0} > 0$. In the cases (ii) and (iii), we know $f(x) \succ x - x_0$ and $x - x_0 \succ g(x)$, and therefore $f(x) \succ g(x)$. It is easy to verify that the h_1, h_2 functions so constructed are constants.

(b) If $g(x) \geq 0$ or $f(x) \leq 0$ for all $x \in \mathbb{R}$, it is trivial. Otherwise, there exists a x_0 such that $g(x_0) < 0$ and $f(x_0) > 0$. Define

$$\hat{f}(x) = (x - x_0)^2 f(\frac{1}{x - x_0})$$

and

$$\hat{g}(x) = (x - x_0)^2 g(\frac{1}{x - x_0}).$$

Because $g(x_0) < 0$, we know $\hat{g}(x)$ is concave quadratic. Similarly, $\hat{f}(x)$ is convex. For all $x \neq x_0$, if $\hat{g}(x) \geq 0$ then $g(\frac{1}{x-x_0}) \geq 0$, therefore $f(\frac{1}{x-x_0}) > 0$, and consequently $\hat{f}(x) > 0$. Also, $\hat{g}(x_0) > 0$ and $\hat{f}(x_0) > 0$ because both f and g are convex quadratic functions.

So we have $\hat{f}|_{\hat{g}\geq 0} > 0$. It follows from (iv.a) that there exist nonnegative constants $\hat{h}_1$, $\hat{h}_2$, and $\hat{h}_3(x) > 0$ for all $x \in \mathbb{R}$ such that

$$\hat{h}_1 \hat{f} = \hat{h}_2 \hat{g} + \hat{h}_3.$$

Let $h_1 = \hat{h}_1$, $h_2 = \hat{h}_2$, $h_3(x) = x^2 \hat{h}_3(\frac{1}{x} + x_0)$. For any $x \neq 0$, let $y = \frac{1}{x} + x_0$. Noticing that $x = \frac{1}{y - x_0}$, we have $f(x) = x^2 \hat{f}(\frac{1}{x} + x_0)$ and $g(x) = x^2 \hat{g}(\frac{1}{x} + x_0)$. Therefore

$$h_1 f(x) - h_2 g(x) - h_3(x) = x^2 \left[\hat{h}_1 \hat{f}(y) - \hat{h}_2 \hat{g}(y) - \hat{h}_3(y) \right] = 0.$$

Taking the limit $x \to 0$, we know that the above equality also holds for $x = 0$. Therefore $h_1 f(x) = h_2 g(x) + h_3(x)$ for all $x \in \mathbb{R}$. For any $x \neq 0$, because $\hat{h}_3(y) > 0$, it follows that $h_3(x) = x^2 \hat{h}_3(y) > 0$. For $x = 0$, $h_3(0) > 0$ follows from the fact that $\hat{h}_3$ is strongly convex (see case (ii)).

$\square$

Lemma 6. *(c.f. Lemma 2) $f_i(x) \succ g(x)$ for $i = 1, \cdots, I \Longrightarrow \prod_{i=1}^{I} f_i(x) \succ g(x)$. The result also holds for the $\succeq$-relation.*

Proof. Consider first $f_1(x) \succ g(x)$ and $f_2(x) \succ g(x)$, i.e. there exist nonnegative $\theta_j^i(x)$ for $i = 1, 2$ and $j = 1, 2, 3$ such that

$$\begin{aligned} &\theta_1^1(x) f_1(x) = \theta_2^1(x) g(x) + \theta_3^1(x) \\ \text{and} \quad &\theta_1^2(x) f_2(x) = \theta_2^2(x) g(x) + \theta_3^2(x) \end{aligned}$$

Then

$$\begin{aligned} \theta_1^1(x)\theta_1^2(x) f_1(x) f_2(x) = &\left(\theta_2^1(x)\theta_2^2(x)[g(x)]^2 + \theta_3^1(x)\theta_3^2(x)\right) \\ &+ \left(\theta_2^1(x)\theta_3^2(x) + \theta_2^2(x)\theta_3^1(x)\right) g(x) \end{aligned}$$

Taking $h_1(x) = \theta_1^1(x)\theta_1^2(x)$, $h_2(x) = \theta_2^1(x)\theta_2^2(x)[g(x)]^2 + \theta_3^1(x)\theta_3^2(x)$ and $h_3(x) = \theta_2^1(x)\theta_3^2(x) + \theta_2^2(x)\theta_3^1(x)$, we prove the claim for $i = 2$.

If we take $\widetilde{f_1}(x) = f_1(x) f_2(x)$ and $\widetilde{f_2}(x) = f_3(x)$, we can repeat the argument to prove the claim for $i = 3$. By induction, the claim is true for $i = 1, \cdots, I$.

The proof for the $\succeq$ case is identical and so is omitted. $\square$

Theorem 6. *(Almost S-Lemma)* $f|_{g \geq 0} > 0 \Longrightarrow f(x) \succ g(x)$.

Proof. We shall prove the theorem in a few steps. First, notice that

$$f|_{g \geq 0} > 0$$

$$\Longrightarrow \exists f_i(x) \text{ with } \deg f_i(x) \leq 2, i = 1, \cdots, I,$$

$$\text{and } f_0(x) > 0 \ \forall x \in \mathbb{R} \text{ such that } f(x) = f_0(x) \prod_{i=1}^{I} f_i(x)$$

$$\text{and } f_i|_{g \geq 0} > 0 \qquad \text{(by Lemma 2)}$$

This further implies that

$$-g|_{-f_i \geq 0} > 0 \text{ for } i = 1, \cdots, I, \qquad \text{(by Lemma 3)}$$

$$\Longrightarrow \forall i = 1, \cdots, I, \exists g_j^i(x) \text{ with } \deg g_j^i(x) \leq 2, j = 1, \cdots, J,$$

$$\text{and } g_0(x) > 0 \ \forall x \in \mathbb{R}$$

$$\text{such that } -g(x) = g_0(x) \prod_{j=1}^{J} g_j^i(x)$$

$$\text{and } g_j^i|_{-f_i \geq 0} > 0 \qquad \text{(by Lemma 2)}$$

$$\Longrightarrow \forall i = 1, \cdots, I, g_j^i(x) \succ -f_i(x),$$

$$j = 1, \cdots, J \qquad \text{(by Lemma 5)}$$

$$\Longrightarrow \forall i = 1, \cdots, I,$$

$$-g(x) = g_0(x) \prod_{j=1}^{J} g_j^i(x) \succ -f_i(x) \qquad \text{(by Lemma 6)}$$

$$\Longleftrightarrow \forall i = 1, \cdots, I, f_i(x) \succ g(x) \qquad \text{(by Lemma 4)}$$

$$\Longrightarrow f(x) = f_0(x) \prod_{i=1}^{I} f_i(x) \succ g(x) \qquad \text{(by Lemma 6)}$$

□

Theorem 7. *Let* $f(x) = f_0(x) \prod_{i=1}^{I} f_i(x)$ *and* $g(x) = g_0(x) \prod_{j=1}^{J} g_j^i(x)$*,* *where* $\deg f_0 := 2r < \deg f$ *and* $\deg g_0 := 2s < \deg g$. *Furthermore, letting* $h_1(x)$ *and* $h_2(x)$ *be defined under the relation* $f(x) \succ g(x)$, *then we have*

1. $\deg h_1(x) = 2s + 2\left\lceil \frac{\deg f(x) - 2r}{2} \right\rceil \left(\left\lceil \frac{\deg g(x) - 2s}{2} \right\rceil - 1 \right)$
2. $\deg h_2(x) = 2r + 2\left\lceil \frac{\deg g(x) - 2s}{2} \right\rceil \left(\left\lceil \frac{\deg f(x) - 2r}{2} \right\rceil - 1 \right)$

Proof. The given decomposition of f and g follow from the proof of Theorem 6, where $\deg f_i(x) \leq 2$ and $\deg g_j^i(x) \leq 2$. Let $\tilde{f}(x) = \prod_{i=1}^{I} f_i(x)$ and $-\tilde{g}(x) = \prod_{j=1}^{J} g_j^i(x)$. We first consider the nonnegative function $\tilde{h}_1(x), \tilde{h}_2(x)$ for $\tilde{f}(x)$ and $\tilde{g}(x)$. Note that $\deg \tilde{f}(x) = \deg f(x) - 2r$ and $\deg \tilde{g}(x) = \deg g(x) - 2s$ From the proof of Lemma 5, there exist constants $\zeta_1^{(i,j)} > 0$,

$\zeta_2^{(i,j)} > 0$ and $\theta^{(i,j)}(x) \geq 0$ with $\deg \theta^{(i,j)}(x) \leq 2$ such that

$$\begin{aligned}
\zeta_1^{(i,j)} g_j^i(x) &= \zeta_2^{(i,j)}(-f_i(x)) + \theta^{(i,j)}(x) \\
\prod_{j=1}^{J} \zeta_1^{(i,j)} g_j^i(x) &= \prod_{j=1}^{J} \left(\zeta_2^{(i,j)}(-f_i(x)) + \theta^{(i,j)}(x)\right) \\
-\tilde{g}(x) \prod_{j=1}^{J} \zeta_1^{(i,j)} &= -\widetilde{\theta}_2^i(x) f_i(x) + \widetilde{\theta}_3^i(x) \quad \text{for some } \widetilde{\theta}_2^i(x) \text{ and } \widetilde{\theta}_3^i(x)
\end{aligned} \tag{4.1}$$

$$\begin{aligned}
\widetilde{\theta}_2^i(x) f_i(x) &= \tilde{g}(x)\tilde{\zeta}^i + \widetilde{\theta}_3^i(x), \text{ where } \tilde{\zeta}^i = \prod_{j=1}^{J} \zeta_1^{(i,j)} \\
\prod_{i=1}^{I} \widetilde{\theta}_2^i(x) f_i(x) &= \prod_{i=1}^{I} \left(\tilde{g}(x)\tilde{\zeta}^i + \widetilde{\theta}_3^i(x)\right) \\
\tilde{h}_1(x)\tilde{f}(x) &= \tilde{h}_2(x)\tilde{g}(x) + \tilde{h}_3(x) \text{ for some } \tilde{h}_2(x) \text{ and } \tilde{h}_3(x)
\end{aligned} \tag{4.2}$$

In (4.1), one can verify that there exists i such that $\deg \widetilde{\theta}_2^i(x) = 2\left(\lceil \frac{\deg g(x) - 2s}{2} \rceil - 1\right)$ and $\deg \widetilde{\theta}_3^i(x) = 2\lceil \frac{\deg g(x)-2s}{2} \rceil$.
Then in (4.2), $\tilde{h}_1(x) = \prod_{i=1}^{I} \widetilde{\theta}_2^i(x)$ and

$$\begin{aligned}
\deg \tilde{h}_1(x) =& \lceil \frac{\deg f(x) - 2r}{2} \rceil \deg \widetilde{\theta}_2^i(x) \\
=& 2\lceil \frac{\deg f(x) - 2r}{2} \rceil \left(\lceil \frac{\deg g(x) - 2s}{2} \rceil - 1\right).
\end{aligned}$$

One can also verify that

$$\begin{aligned}
\deg \tilde{h}_2(x) =& \lceil \frac{\deg g(x) - 2s}{2} \rceil \deg \widetilde{\theta}_2^i(x) \\
=& 2\lceil \frac{\deg g(x) - 2s}{2} \rceil \left(\lceil \frac{\deg f(x) - 2r}{2} \rceil - 1\right).
\end{aligned}$$

Multiplying $f_0(x)g_0(x)$ on both sides of (4.2), we recover $f(x)$ and $g(x)$ with $h_1(x) = g_0(x)\tilde{h}_1(x)$ and $h_2(x) = f_0(x)\tilde{h}_2(x)$ with their respective claimed degree. $\square$

Let us remark that Theorem 7 does not apply and it is a trivial case when $\deg \tilde{f}(x) = \deg f(x)$ or $\deg \tilde{g}(x) = \deg g(x)$. Meanwhile, we have a concern about the knowledge of r and s in Theorem 7. Given $g(x)$ in a constraint of robust optimization, coefficients of $f(x)$ are unknown and so is r. Fortunately this can be resolved when we can make a practical assumption that $\deg g(x) \leq 2$. In other words, we can arrive at a bound of $\deg h_2(x)$ without r under the assumption:

Corollary 1. *Assume* $\deg g(x) \leq 2$. *Then*

$$\deg h_2(x) = 2\left(\left\lceil \frac{\deg f(x)}{2} \right\rceil - 1\right).$$

This is equivalent to the case when $r = 0$ *in Theorem 7.*

Proof. $\deg g(x) \leq 2$ implies $s = 0$ and $\left\lceil \frac{\deg g(x) - 2s}{2} \right\rceil$=1. Note that $\left\lceil \frac{\deg f(x) - 2r}{2} \right\rceil = \left\lceil \frac{\deg f(x)}{2} \right\rceil - r$. Then according to Theorem 7,

$$\begin{aligned} \deg h_2(x) =& 2r + 2\left(\left\lceil \frac{\deg f(x)}{2} \right\rceil - r - 1\right) \\ =& 2\left(\left\lceil \frac{\deg f(x)}{2} \right\rceil - 1\right) \end{aligned}$$

$\square$

The construction of the functions h_1 and h_2 is presented in the proof of Theorem 7. Let us illustrate the theorems together with the constructions with the following examples.

Example 5. *Let*

$$f(x) = x$$
$$\text{and } g(x) = (x-1)(x-2)(x-3).$$

We have $f|_{g\geq 0} \geq 0$.
By Lemma 3, $-g|_{-f\geq 0} \geq 0$.
Now, decompose $-g(x) = g_1(x)g_2(x)$*, where* $g_1(x) = (x-1)(x-2)$ *and* $g_2(x) = -(x-3)$.
By Lemma 3 again, we have

$$g_1|_{-f\geq 0} \geq 0, \quad g_2|_{-f\geq 0} \geq 0$$

Then by Lemma 5, construct $\theta_3^j(x)$*, for j=1,2, such that*

$$\begin{aligned} \theta_3^1(x) = & \; g_1(x) - (-f_1(x)) & = x^2 - 2x + 2 \\ \theta_3^2(x) = & \; g_2(x) - (-f_1(x)) & = 3 \end{aligned}$$

These imply

$$\begin{aligned} g_1(x) \cdot g_2(x) &= \left[\theta_3^1(x) + (-f(x))\right]\left[\theta_3^2(x) + (-f(x))\right] \\ -g(x) &= -(x^2 - 2x + 5)f(x) + (4x^2 - 6x + 6) \end{aligned}$$

$$(x^2 - 2x + 5)f(x) = g(x) + (4x^2 - 6x + 6)$$

Here

$$h_1(x) = x^2 - 2x + 5$$
$$h_2(x) = 1$$

We can verify that

$$\begin{aligned}
&2\left\lceil\frac{\deg f(x)}{2}\right\rceil\left(\left\lceil\frac{\deg g(x)}{2}\right\rceil - 1\right) \\
&= 2 \cdot \left\lceil\frac{1}{2}\right\rceil \cdot (\left\lceil\frac{3}{2}\right\rceil - 1) = 2 = \deg h_1(x) \\
&2\left\lceil\frac{\deg g(x)}{2}\right\rceil\left(\left\lceil\frac{\deg f(x)}{2}\right\rceil - 1\right) \\
&= 2 \cdot \left\lceil\frac{3}{2}\right\rceil \cdot (\left\lceil\frac{1}{2}\right\rceil - 1) = 0 = \deg h_2(x)
\end{aligned}$$

Example 6. *Let*

$$f(x) = (x+1)(x-3)(x-5)$$
$$\textit{and } g(x) = x(x-1)(x-6)(x-8)(x-9).$$

We have $f|_{g\geq 0} \geq 0$.

Decompose $f(x) = f_1(x)f_2(x)$, *where* $f_1(x) = x+1$ *and* $f_2(x) = (x-3)(x-5)$. *Note that* $f_1|_{g\geq 0} \geq 0$ *and* $f_2|_{g\geq 0} \geq 0$.

By Lemma 3, $-g|_{-f_1\geq 0} \geq 0$ *and* $-g|_{-f_2\geq 0} \geq 0$.

Now, decompose $-g(x) = g_1(x)g_2(x)g_3(x)$, *where* $g_1(x) = x(x-1)$, $g_2(x) = (x-6)(x-8)$ *and* $g_3(x) = -(x-9)$.

By Lemma 3 again, we have

$$g_1|_{-f_1\geq 0} \geq 0, \quad g_2|_{-f_1\geq 0} \geq 0, \quad g_3|_{-f_1\geq 0} \geq 0,$$
$$g_1|_{-f_2\geq 0} \geq 0, \quad g_2|_{-f_2\geq 0} \geq 0, \quad g_3|_{-f_2\geq 0} \geq 0.$$

Then by Lemma 5, construct $\theta_3^{(i,j)}(x)$, *for* $i = 1, 2$ *and j=1,2,3, such that*

$$\begin{aligned}
\theta_3^{(1,1)}(x) &= \quad g_1(x) - (-f_1(x)) &&= x^2 + 1 \\
\theta_3^{(1,2)}(x) &= \quad \tfrac{1}{2}g_2(x) - (-f_1(x)) &&= \frac{1}{2}x^2 - 6x + 25 \\
\theta_3^{(1,3)}(x) &= \quad g_3(x) - (-f_1(x)) &&= 10 \\
\theta_3^{(2,1)}(x) &= \quad g_1(x) - \tfrac{1}{2}(-f_2(x)) &&= \frac{3}{2}x^2 - 5x + \frac{15}{2} \\
\theta_3^{(2,2)}(x) &= \quad \tfrac{1}{2}g_2(x) - \tfrac{1}{2}(-f_2(x)) &&= x^2 - 11x + \frac{63}{2} \\
\theta_3^{(2,3)}(x) &= \quad g_3(x) - (-f_2(x)) &&= x^2 - 9x + 24
\end{aligned}$$

The first three expressions and last three imply, respectively,

$$\begin{aligned}
g_1(x) \cdot \frac{1}{2} g_2(x) \cdot g_3(x) = & \left[\theta_3^{(1,1)}(x) + (-f_1(x))\right] \cdot \\
& \left[\theta_3^{(1,2)}(x) + (-f_1(x))\right] \cdot \\
& \left[\theta_3^{(1,3)}(x) + (-f_1(x))\right], \\
\frac{1}{2}[-g(x)] = & -\widetilde{\theta}_2^1(x) f_1(x) + \widetilde{\theta}_3^1(x), && (4.3) \\
\text{and } g_1(x) \cdot \frac{1}{2} g_2(x) \cdot g_3(x) = & \left[\theta_3^{(2,1)}(x) + \frac{1}{2}(-f_2(x))\right] \cdot \\
& \left[\theta_3^{(2,2)}(x) + \frac{1}{2}(-f_2(x))\right] \cdot \\
& \left[\theta_3^{(2,3)}(x) + (-f_2(x))\right], \\
\frac{1}{2}[-g(x)] = & -\widetilde{\theta}_2^2(x) f_2(x) + \widetilde{\theta}_3^2(x), && (4.4)
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\theta}_2^1(x) = & \left[\theta_3^{(1,1)}(x)\theta_3^{(1,2)}(x) + (-f_1(x))^2\right] \\
& + \theta_3^{(1,3)}(x)\left[\theta_3^{(1,2)}(x) + \theta_3^{(1,1)}(x)\right] \\
= & \frac{x^4}{2} - 6x^3 + \frac{83x^2}{2} - 64x + 286
\end{aligned}$$

$$\begin{aligned}
\widetilde{\theta}_3^1(x) = & \left[\theta_3^{(1,1)}(x)\theta_3^{(1,2)}(x) + (-f_1(x))^2\right]\theta_3^{(1,3)}(x) \\
& + \left[\theta_3^{(1,2)}(x) + \theta_3^{(1,1)}(x)\right](-f_1(x))^2 \\
= & \frac{13x^4}{2} - 63x^3 + \frac{561x^2}{2} + 6x + 286
\end{aligned}$$

$$\begin{aligned}\widetilde{\theta}_2^2(x) = &\left[\theta_3^{(2,1)}(x)\theta_3^{(2,2)}(x) + \frac{1}{4}(-f_2(x))^2\right] \\ &+ \theta_3^{(2,3)}(x)\left[\frac{1}{2}\theta_3^{(2,1)}(x) + \frac{1}{2}\theta_3^{(2,2)}(x)\right] \\ =&3x^4 - \frac{179x^3}{4} + \frac{1019x^2}{4} - \frac{1335x}{2} + \frac{1521}{2}\end{aligned}$$

$$\begin{aligned}\widetilde{\theta}_3^2(x) = &\left[\theta_3^{(2,1)}(x)\theta_3^{(2,2)}(x) + \frac{1}{4}(-f_2(x))^2\right]\theta_3^{(2,3)}(x) \\ &+ \left[\frac{1}{2}\theta_3^{(2,1)}(x) + \frac{1}{2}\theta_3^{(2,2)}(x)\right](-f_2(x))^2 \\ &= 3x^6 - \frac{277x^5}{4} + \frac{2679x^4}{4} - \frac{13901x^3}{4} \\ &+ \frac{40899x^2}{4} - \frac{32625x}{2} + \frac{22815}{2}\end{aligned}$$

Then from (4.3) and (4.4)

$$\widetilde{\theta}_2^1(x)f_1(x) \cdot \widetilde{\theta}_2^2(x)f_2(x) = \left[\frac{1}{2}g(x) + \widetilde{\theta}_3^1(x)\right]\left[\frac{1}{2}g(x) + \widetilde{\theta}_3^2(x)\right]$$

$$h_1(x)f(x) = h_2(x)g(x) + h_3(x)$$

where

$$\begin{aligned}
h_1(x) =& \widetilde{\theta}_2^1(x)\widetilde{\theta}_2^2(x)\\
=& \frac{3x^8}{2} - \frac{323x^7}{8} + \frac{4163x^6}{8} - \frac{31291x^5}{8} + \frac{149435x^4}{8}\\
& - \frac{245467x^3}{4} + \frac{588557x^2}{4} - 239577x + 217503,\\
h_2(x) =& \frac{1}{2}\left[\widetilde{\theta}_3^1(x) + \widetilde{\theta}_3^2(x)\right]\\
=& \frac{3x^6}{2} - \frac{277x^5}{8} + \frac{2705x^4}{8} - \frac{14153x^3}{8}\\
& + \frac{42021x^2}{8} - \frac{32613x}{4} + \frac{23387}{4}
\end{aligned}$$

We can verify that

$$\begin{aligned}
& 2\left\lceil\frac{\deg f(x)}{2}\right\rceil\left(\left\lceil\frac{\deg g(x)}{2}\right\rceil - 1\right)\\
& = 2\cdot\left\lceil\frac{3}{2}\right\rceil\cdot(\left\lceil\frac{5}{2}\right\rceil - 1) = 8 = \deg h_1(x),\\
& 2\left\lceil\frac{\deg g(x)}{2}\right\rceil\left(\left\lceil\frac{\deg f(x)}{2}\right\rceil - 1\right)\\
& = 2\cdot\left\lceil\frac{5}{2}\right\rceil\cdot(\left\lceil\frac{3}{2}\right\rceil - 1) = 6 = \deg h_2(x)
\end{aligned}$$

Example 7. *Let*

$$f(x) = (x-1)(x-3)(x-6)(x-7)$$
$$\text{and } g(x) = x(x-4)(x-5)(x-8).$$

We have $f|_{g\geq 0} \geq 0$.

Decompose $f(x) = f_1(x)f_2(x)$, *where* $f_1(x) = (x-1)(x-3)$ *and*

$f_2(x) = (x-6)(x-7)$. *Note that* $f_1|_{g\geq 0} \geq 0$ *and* $f_2|_{g\geq 0} \geq 0$.
By Lemma 3, $-g|_{-f_1\geq 0} \geq 0$ *and* $-g|_{-f_2\geq 0} \geq 0$.
Now, decompose $-g(x) = g_1(x)g_2(x)$, *where* $g_1(x) = (x-4)(x-5)$ *and* $g_2(x) = -x(x-8)$.
By Lemma 3 again, we have

$$g_1|_{-f_1\geq 0} \geq 0, \quad g_2|_{-f_1\geq 0} \geq 0,$$
$$g_1|_{-f_2\geq 0} \geq 0, \quad g_2|_{-f_2\geq 0} \geq 0.$$

Then by Lemma 5, construct $\theta_3^{(i,j)}(x)$, *for* $i = 1,2$ *and j=1,2, such that*

$$\begin{aligned}
\theta_3^{(1,1)}(x) = &\quad g_1(x) - \tfrac{1}{2}(-f_1(x)) &&= \frac{3}{2}x^2 - 11x + \frac{43}{2} \\
\theta_3^{(1,2)}(x) = &\quad \tfrac{1}{64}g_2(x) - \tfrac{1}{16}(-f_1(x)) &&= \frac{1}{64}(3x^2 - 8x + 12) \\
\theta_3^{(2,1)}(x) = &\quad g_1(x) - (-f_2(x)) &&= 2x^2 - 22x + 62 \\
\theta_3^{(2,2)}(x) = &\quad \tfrac{1}{64}g_2(x) - \tfrac{1}{8}(-f_2(x)) &&= \frac{1}{64}(7x^2 - 96x + 336)
\end{aligned}$$

The first two expressions and last two imply, respectively,

$$g_1(x) \cdot \frac{1}{64}g_2(x) = \left[\theta_3^{(1,1)}(x) + \frac{1}{2}(-f_1(x))\right]\left[\theta_3^{(1,2)}(x) + \frac{1}{16}(-f_1(x))\right]$$
$$\frac{1}{64}[-g(x)] = -\widetilde{\theta}_2^1(x)f_1(x) + \widetilde{\theta}_3^1(x), \tag{4.5}$$
and $$g_1(x) \cdot \frac{1}{64}g_2(x) = \left[\theta_3^{(2,1)}(x) + (-f_2(x))\right]\left[\theta_3^{(2,2)}(x) + \frac{1}{8}(-f_2(x))\right]$$
$$\frac{1}{64}[-g(x)] = -\widetilde{\theta}_2^2(x)f_2(x) + \widetilde{\theta}_3^2(x), \tag{4.6}$$

where

$$\begin{aligned}
\widetilde{\theta}_2^1(x) &= \left[\frac{1}{16}\theta_3^{(1,1)}(x) + \frac{1}{2}\theta_3^{(1,2)}(x)\right] \\
&= \frac{15x^2}{128} - \frac{3x}{4} + \frac{23}{16} \\
\widetilde{\theta}_3^1(x) &= \left[\theta_3^{(1,1)}(x)\theta_3^{(1,2)}(x) + \frac{1}{32}(-f_1(x))^2\right] \\
&= \frac{13x^4}{128} - \frac{61x^3}{64} + \frac{429x^2}{128} - \frac{11x}{2} + \frac{69}{16} \\
\widetilde{\theta}_2^2(x) &= \left[\frac{1}{8}\theta_3^{(2,1)}(x) + \theta_3^{(2,2)}(x)\right] \\
&= \frac{23x^2}{64} - \frac{17x}{4} + 13 \\
\widetilde{\theta}_3^2(x) &= \left[\theta_3^{(2,1)}(x)\theta_3^{(2,2)}(x) + \frac{1}{8}(-f_2(x))^2\right] \\
&= \frac{11x^4}{32} - \frac{277x^3}{32} + \frac{2621x^2}{32} - 345x + 546
\end{aligned}$$

Then from (4.3) and (4.4)

$$\widetilde{\theta}_2^1(x)f_1(x) \cdot \widetilde{\theta}_2^2(x)f_2(x) = \left[\frac{1}{64}g(x) + \widetilde{\theta}_3^1(x)\right]\left[\frac{1}{64}g(x) + \widetilde{\theta}_3^2(x)\right]$$

$$h_1(x)f(x) = h_2(x)g(x) + h_3(x)$$

where

$$\begin{aligned}
h_1(x) =& \widetilde{\theta}_2^1(x)\widetilde{\theta}_2^2(x) \\
=& \frac{345x^4}{8192} - \frac{393x^3}{512} + \frac{5353x^2}{1024} - \frac{1015x}{64} + \frac{299}{16} \\
h_2(x) =& \frac{1}{64}\left[\widetilde{\theta}_3^1(x) + \widetilde{\theta}_3^2(x)\right] \\
=& \frac{57x^4}{8192} - \frac{615x^3}{4096} + \frac{10913x^2}{8192} - \frac{701x}{128} + \frac{8805}{1024}
\end{aligned}$$

We can verify that

$$2\left\lceil\frac{\deg f(x)}{2}\right\rceil\left(\left\lceil\frac{\deg g(x)}{2}\right\rceil-1\right)=2\cdot\frac{4}{2}\cdot(\frac{4}{2}-1)=4=\deg h_1(x)$$
$$2\left\lceil\frac{\deg g(x)}{2}\right\rceil\left(\left\lceil\frac{\deg f(x)}{2}\right\rceil-1\right)=2\cdot\frac{4}{2}\cdot(\frac{4}{2}-1)=4=\deg h_2(x)$$

Example 8. *Let*

$$f(x) = x(x-1)(x-6)(x-7)$$
$$\text{and } g(x) = (x+1)(x-3)(x-5)(x-8)(x-9)(x-10).$$

We have $f|_{g\geq 0} \geq 0$.

Decompose $f(x) = f_1(x)f_2(x)$, *where* $f_1(x) = x(x-1)$ *and* $f_2(x) = (x-6)(x-7)$. *Note that* $f_1|_{g\geq 0} \geq 0$ *and* $f_2|_{g\geq 0} \geq 0$.

By Lemma 3, $-g|_{-f_1\geq 0} \geq 0$ *and* $-g|_{-f_2\geq 0} \geq 0$.

Now, decompose $-g(x) = g_1(x)g_2(x)g_3(x)$, *where* $g_1(x) = -(x+1)(x-10)$, $g_2(x) = (x-3)(x-5)$ *and* $g_3(x) = (x-8)(x-9)$.

By Lemma 3 again, we have

$$g_1|_{-f_1\geq 0} \geq 0, \quad g_2|_{-f_1\geq 0} \geq 0, \quad g_3|_{-f_1\geq 0} \geq 0,$$
$$g_1|_{-f_2\geq 0} \geq 0, \quad g_2|_{-f_2\geq 0} \geq 0, \quad g_3|_{-f_2\geq 0} \geq 0.$$

Then by Lemma 5, construct $\theta_3^{(i,j)}(x)$, *for* $i = 1, 2$ *and j=1,2,3,* *such that*

$$\begin{aligned}
\theta_3^{(1,1)}(x) = &\ \tfrac{1}{11^2} g_1(x) - \tfrac{1}{11}(-f_1(x)) &&= \frac{2}{11^2}(5x^2 - x + 5) \\
\theta_3^{(1,2)}(x) = &\ \tfrac{1}{2} g_2(x) - (-f_1(x)) &&= \frac{3}{2}x^2 - 5x + \frac{15}{2} \\
\theta_3^{(1,3)}(x) = &\ g_3(x) - (-f_1(x)) &&= 2x^2 - 18x + 72 \\
\theta_3^{(2,1)}(x) = &\ \tfrac{1}{11^2} g_1(x) - \tfrac{1}{11}(-f_2(x)) &&= \frac{2}{11^2}(5x^2 - 67x + 236) \\
\theta_3^{(2,2)}(x) = &\ \tfrac{1}{2} g_2(x) - (-f_2(x)) &&= \frac{3}{2}x^2 - 17x + \frac{99}{2} \\
\theta_3^{(2,3)}(x) = &\ g_3(x) - (-f_2(x)) &&= 2x^2 - 30x + 114
\end{aligned}$$

The first three expressions and last three imply, respectively,

$$\begin{aligned}
\frac{1}{11^2} g_1(x) \cdot \frac{1}{2} g_2(x) \cdot g_3(x) = &\left[\theta_3^{(1,1)}(x) + \frac{1}{11}(-f_1(x))\right] \cdot \\
&\left[\theta_3^{(1,2)}(x) + (-f_1(x))\right] \cdot \\
&\left[\theta_3^{(1,3)}(x) + (-f_1(x))\right], \\
\frac{1}{242}[-g(x)] = &- \widetilde{\theta}_2^1(x) f_1(x) + \widetilde{\theta}_3^1(x), \qquad (4.7)
\end{aligned}$$

$$\begin{aligned}
\text{and } \frac{1}{11^2} g_1(x) \cdot \frac{1}{2} g_2(x) \cdot g_3(x) = &\left[\theta_3^{(2,1)}(x) + \frac{1}{11}(-f_2(x))\right] \cdot \\
&\left[\theta_3^{(2,2)}(x) + (-f_2(x))\right] \cdot \\
&\left[\theta_3^{(2,3)}(x) + (-f_2(x))\right], \\
\frac{1}{242}[-g(x)] = &- \widetilde{\theta}_2^2(x) f_2(x) + \widetilde{\theta}_3^2(x), \qquad (4.8)
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\theta}_2^1(x) = &\left[\theta_3^{(1,1)}(x)\theta_3^{(1,2)}(x) + \frac{1}{11}(-f_1(x))^2\right] \\
&+ \theta_3^{(1,3)}(x)\left[\frac{1}{11}\theta_3^{(1,2)}(x) + \theta_3^{(1,1)}(x)\right] \\
=&\frac{79x^4}{121} - \frac{666x^3}{121} + \frac{3230x^2}{121} - \frac{5834x}{121} + \frac{6735}{121}
\end{aligned}$$

$$\begin{aligned}
\widetilde{\theta}_3^1(x) = &\left[\theta_3^{(1,1)}(x)\theta_3^{(1,2)}(x) + \frac{1}{11}(-f_1(x))^2\right]\theta_3^{(1,3)}(x) \\
&+ \left[\frac{1}{11}\theta_3^{(1,2)}(x) + \theta_3^{(1,1)}(x)\right](-f_1(x))^2 \\
=&\frac{157x^6}{242} - \frac{728x^5}{121} + \frac{3677x^4}{121} - \frac{7770x^3}{121} \\
&+ \frac{18809x^2}{242} - \frac{6030x}{121} + \frac{5400}{121}
\end{aligned}$$

$$\begin{aligned}
\widetilde{\theta}_2^2(x) = &\left[\theta_3^{(2,1)}(x)\theta_3^{(2,2)}(x) + \frac{1}{11}(-f_2(x))^2\right] \\
&+ \theta_3^{(2,3)}(x)\left[\frac{1}{11}\theta_3^{(2,2)}(x) + \theta_3^{(2,1)}(x)\right] \\
=&\frac{79x^4}{121} - \frac{2094x^3}{121} + \frac{20948x^2}{121} - \frac{93758x}{121} + \frac{158649}{121}
\end{aligned}$$

$$\begin{aligned}
\widetilde{\theta}_3^2(x) = &\left[\theta_3^{(2,1)}(x)\theta_3^{(2,2)}(x) + \frac{1}{11}(-f_2(x))^2\right]\theta_3^{(2,3)}(x) \\
&+ \left[\frac{1}{11}\theta_3^{(2,2)}(x) + \theta_3^{(2,1)}(x)\right](-f_2(x))^2 \\
=&\frac{157x^6}{242} - \frac{3104x^5}{121} + \frac{51269x^4}{121} - \frac{452736x^3}{121} \\
&+ \frac{4508309x^2}{242} - \frac{5999568x}{121} + \frac{6668658}{121}
\end{aligned}$$

Then from (4.7) and (4.8)

$$\widetilde{\theta}_2^1(x)f_1(x) \cdot \widetilde{\theta}_2^2(x)f_2(x) = \left[\frac{1}{242}g(x) + \widetilde{\theta}_3^1(x)\right]\left[\frac{1}{242}g(x) + \widetilde{\theta}_3^2(x)\right]$$

$$h_1(x)f(x) = h_2(x)g(x) + h_3(x)$$

where

$$\begin{aligned}
h_1(x) =& \widetilde{\theta}_2^1(x)\widetilde{\theta}_2^2(x) \\
=& \frac{6241x^8}{14641} - \frac{218040x^7}{14641} + \frac{3304666x^6}{14641} - \frac{28582756x^5}{14641} \\
&+ \frac{155386600x^4}{14641} - \frac{544812296x^3}{14641} + \frac{1200505222x^2}{14641} \\
&- \frac{1557018396x}{14641} + \frac{1068501015}{14641} \\
h_2(x) =& \frac{1}{242}\left[\widetilde{\theta}_3^1(x) + \widetilde{\theta}_3^2(x)\right] \\
=& \frac{157x^6}{29282} - \frac{1916x^5}{14641} + \frac{27473x^4}{14641} - \frac{230253x^3}{14641} \\
&+ \frac{2263559x^2}{29282} - \frac{3002799x}{14641} + \frac{3337029}{14641}
\end{aligned}$$

We can verify that

$$2\left\lceil\frac{\deg f(x)}{2}\right\rceil\left(\left\lceil\frac{\deg g(x)}{2}\right\rceil - 1\right) = 2 \cdot \frac{4}{2} \cdot (\frac{6}{2} - 1) = 8 = \deg h_1(x)$$
$$2\left\lceil\frac{\deg g(x)}{2}\right\rceil\left(\left\lceil\frac{\deg f(x)}{2}\right\rceil - 1\right) = 2 \cdot \frac{6}{2} \cdot (\frac{4}{2} - 1) = 6 = \deg h_2(x)$$

Theorem 8. *(S-Lemma) Assume* $\gcd(g, g') = 1$ *(regularity condition). Then* $f|_{g\geq 0} \geq 0 \iff f(x) \succeq g(x)$

Proof. :

"$\Longrightarrow$ " The arguments in the proof in Theorem 6 remains valid for $f|_{g\geq 0} \geq 0$.

"$\Longleftarrow$ " We are going to show a contradiction for "$f(x) < 0$ when $g(x) \geq 0$", given that $\gcd(g, g') = 1$.

Note that $\gcd(g, g') = 1$ implies that if x^* is a root of g, then x^* must not be a local maximum.

On the other hand, consider $h_1 f = h_2 g + h_3$, as defined in $\succeq$. $g(x^*) \geq 0 \Rightarrow h_1(x^*)f(x^*) \geq 0$. If $f(x^*) < 0$, then $h_1(x^*) = 0$. By continuity of f and h_1, $\exists\ \epsilon > 0$ such that $\forall\ \bar{x} \in [x^* - \epsilon, x^* + \epsilon]$, we have $f(\bar{x}) < 0$ and $h_1(\bar{x}) > 0$. Then $h_2(\bar{x})g(\bar{x}) = f(\bar{x})h_1(\bar{x}) - h_3(\bar{x}) < 0$, or $g(\bar{x}) < 0 \forall\ \bar{x} \in [x^* - \epsilon, x^* + \epsilon]$. Together with the fact that $g(x^*) = 0$, x^* is a local maximum of g, contradicting the assumption $\gcd(g, g') = 1$.

$\square$

S-Lemma may not hold if we remove the regularity condition. Please refer to the following example.

Example 9. *Let* $f(x) = -1$, $g(x) = -x^2$. *Note that* $\gcd(g, g') = x$. *We can choose* $h_1(x) = x^2$, $h_2(x) = 1$ *and* $h_3(x) = 0$. *Then we have* $f(x) \succeq g(x)$, *but* $f|_{g\geq 0} \not\geq 0$.

Two theorems on nonnegative polynomials can be deduced from our S-Lemma.

Corollary 2. *(Polya and Szego [58]) If a polynomial $p(x)$ that satisfies $p(x) \geq 0\ \forall\, x \geq 0$, then there exists nonnegative polynomials $\theta_1(x)$ and $\theta_2(x)$ such that*

$$p(x) = \theta_1(x) + x\theta_2(x) \qquad \forall\, x \in \mathbf{R}$$

Proof. Take $g(x) = x$ and verify that $\gcd(g, g') = 1$. S-Lemma can then apply. $\square$

Corollary 3. *(Fekete (1935), see Lasserre [39], Theorem 2.7 and page 49) The polynomial $p(x)$ satisfies $p(x) \geq 0\ \forall\, x \in [0, 1]$ if and only if there exists nonnegative polynomials $\theta_1(x)$ and $\theta_2(x)$ such that*

$$p(x) = \theta_1(x) + x(1-x)\theta_2(x) \qquad \forall\, x \in \mathbf{R}$$

Proof. Take $g(x) = -x(x-1)$ and verify that $\gcd(g, g') = 1$. S-Lemma can then apply. $\square$

For the original S-Lemma, there is a useful result from Yuan [82] regarding the nonnegativity of the maximum of two quadratic functions. It turns out that, for polynomials, this nonnegativity associates with our version of the S-Lemma as well.

Corollary 4. *(Lemma 2.3 of Yuan [82], or Theorem 5.1 of Ai et al. [1].) Let $f(x)$ and $g(x)$ be two polynomials.*

$$\max\{f(x), g(x)\} \geq 0 \; \forall \, x \in \mathbf{R} \iff f|_{-g>0} \geq 0$$

Furthermore, if $gcd(g, g') = 1$, then there exist $h_1(x) > 0$ and $h_2(x) \geq 0$ such that

$$h_1(x)f(x) + h_2(x)g(x) \geq 0 \quad \forall x \in \mathbf{R} \tag{4.9}$$

Proof.

$$\begin{aligned} & \max\{f(x), g(x)\} \geq 0 \quad \forall \, x \in \mathbf{R} \\ \iff & \begin{cases} g(x) < 0 \Rightarrow f(x) \geq 0 \\ f(x) < 0 \Rightarrow g(x) \geq 0 \end{cases} \\ \iff & f|_{-g>0} \geq 0 \end{aligned}$$

Note that the two statements in the second step are contrapositive of each other, and therefore they are reduced to one statement. By Theorem 8, (4.9) holds. □

Remark. *When the regularity condition $gcd(g, g') = 1$ is imposed, $f|_{-g>0} \geq 0$ is equivalent to $f|_{-g\geq 0} \geq 0$. Otherwise, consider $f(x) = x^2 - 1$ and $g(x) = -x^2$. $f|_{-g>0} \geq 0$ is true while $f|_{-g\geq 0} \geq 0$ is not.*

□ **End of chapter.**

Chapter 5

Confidence bands in polynomial regression

5.1 An introduction

Simultaneous confidence bands are statistical tools to quantify unknown functions. A well-known example is the confidence band for a cumulative distribution function based on the Kolmogorov-Smirnov statistic (see Frey [22] and the references therein for recent advances). Recently, Liu [45] provides a comprehensive overview of the construction methods of simultaneous confidence bands for linear regression models. Based on our results in the previous chapter, we introduce a new computation methodology of the confidence bands for polynomial regression models, which of course includes the linear regression as a special instance.

Robust optimization, together with its SDP formulation(s), is a recent active area of research in Operations Research. While the former is appealing to worst-case modeling, SDP can be computed efficiently through well developed algorithms (e.g. interior point algorithms). This combination of modeling techniques has shown significant applications in various fields, including inventory management, financial engineering and signal processing. Application on the simultaneous confidence bands in polynomial regression is another good example. On the other hand, simultaneous confidence bands in regression analysis play a significant role in various research areas, ranging from econometrics to biostatistics. For instance, in time series analysis, autoregression is studied on unemployment rate against various univariate series; in cardiology, QT/QTc study, which involves the comparison of the regressions and confidence bands analysis, is carried out to access a new drug for possible prolongation of the cardiac repolarization time, which is a biomarker for a potential life-threatening arrhythmia. While there is an existing method to compute the confidence bands, we propose the new approach (through SDP) for the issue.

5.1.1 A review on robust optimization, nonnegative polynomials and SDP

Robust optimization (see Bel-Tal et al. [2] and Nemirovski [54, 55]) is a very large topic in itself, since the idea of including uncertainties in decision making is applicable almost everywhere. It is most practical when the resulting model is tractable (see Ben-Tal et al. [3] and Ben-Tal and Nemirovski [5]), which often refers to the cases that can be modeled or approximated by SDPs. To establish the relationship between the robustness and the confidence bands, we will limit our discussion to nonnegative polynomials and its SDP formulations.

A nonnegative polynomial refers to the nonnegativity of the polynomial over a set. Sum-of-squares (sos) polynomials (see the definition below) have a very close relationship with nonnegative polynomials.

Definition 3. *A polynomial $p(x)$ is sum-of-squares (sos) if and only if there exists polynomials $g_i(x)$, $i = 1, \cdots, m$, such that $p(x) = \sum_{i=1}^{m}[g(x)]^2$. ($x \in \mathbf{R}^n$)*

Hilbert in 1888 identified three cases where nonnegative polynomials and sos polynomials are equivalent: (i) univariate polynomials; (ii) multivariate quadratic polynomials; and (iii) bivariate quartic polynomials. In general, the equivalence is not guaranteed. A classical counterexample that is nonnegative but not sos is given by Motzkin:

$$M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$$

Indeed, checking the nonnegativity of a given polynomial is NP hard, while checking whether it is sos or not is easy: it reduces to solving an SDP:

Lemma 7. *A polynomial $p(x) = \sum_{i=0}^{2n} p_{2n-i}x^i$ is sos if and only if there exists a semidefinite matrix Q such that $p_{2n-k} = \sum_{i+j=k+2} Q_{(i,j)}$, for $i = 0, \cdots, 2n$.*

Together with Hilbert's case (i), we can formulate an SDP for univariate polynomials. This will be further elaborated in the preliminary section.

5.1.2 A review on the confidence bands

A confidence band provides a possible range of true (though unknown) regression curves: An estimated regression curve is plausible if and only if it lies completely inside this range. The earliest discussion on confidence bands, which was considered over the whole real line, was due to Working and Hotelling [77], while Wynn and Bloom [79] and Uusipaikka [75] discussed the construction over a finite interval. Recently, Liu et al. [46]

reviewed the construction of exact confidence bands in linear regression. For quadratic regression, the construction were discussed in Wynn and Bloom [79]and Spurrier [68]. For a general degree polynomial regression, Wynn [78] constructed a piecewise polynomial band. Naiman [52] gave a more conservative band for a more general regression model, in which polynomial regression is applicable. Liu et al. [47] proposed a simulation approach to construct confidence bands for polynomial regression. Most recently, Liu [45] gave a most comprehensive picture of confidence bands in various regression analysis and their applications.

5.1.3 Our contribution

We propose a simulation approach to the construction of confidence bands for polynomial regression of a general degree, which is new in this context and is completely different from Liu et al. [47]. We can generalize the construction over the whole real line, a semi-infinite interval and a bounded interval in a neat and unified approach. In fact, the approach can also be extended to construct the confidence bands over a disjoint union of intervals.

Given the knowledge on nonnegative polynomials and their associations with optimization, we propose a simulation scheme that is fundamentally backed by a single theorem: univariate S-Lemma, which is developed in the previous chapter. (See also Nesterov [56]). We will show how the construction of confidence bands is modeled by semidefinite programming (SDP). There have been software toolboxes of Matlab like cvx and sostool to support the computation of the SDP models, and we will show our numerical results in comparison to those in Liu et al. [47].

In Section 5.2, we briefly introduce some preliminary knowledge on robust optimization, SDP and nonnegative polynomials. In Section 5.3, there are some review on linear regression and confidence region. In Section 5.4, the optimization approach to the computation of confidence bands is presented. Some numerical experiment are conducted in Section 5.5, followed by a conclusion in Section 5.6.

5.2 Some preliminaries on optimization

5.2.1 Robust optimization

Robust optimization is a modeling tool for uncertainties, where the uncertainties (or robustness) are often sought to parameter inaccuracy. Consider a linear programming:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

In the problem, x is the decision model, while $[c, A, b]$ is the data input which is subject to inaccuracy. If we assume a uncertain set $\mathcal{U}$ to include all the realizations of $[c, A, b]$, it is natural to qualify the quality of the solution x by looking into the objective value under the worst-case: $\sup\{c^T x : [c, A, b] \in \mathcal{U}\}$. As a consequence, the robust counterpart of the linear program is

$$\min_{x,t}\{t : c^T x \leq t, Ax \leq b \; \forall [c, A, b] \in \mathcal{U}\}$$

For more details, interested readers may refer to Ben-Tal [2]. Robust optimization may not be computationally tractable in general. However, with certain assumptions on $\mathcal{U}$, it can be formulated as a linear, second-order cone, or semidefinite programming, which can be efficiently solved.

5.2.2 Semidefinite programming and LMIs

Semidefinite Programming is an optimization problem with a linear objective function over the intersection of the cone of positive semidefinite matrices (in a dimension n by n) $\mathcal{S}^n_+$ with an affine space, typically appearing in a primal form:

$$\begin{aligned} \inf \quad & C \bullet X \\ \text{s.t.} \quad & A_i \bullet X \leq b_i \quad \forall i = 1, \cdots, n \\ & X \succeq 0 \end{aligned}$$

where $A \bullet B := tr(A^T B)$ and $X \succeq 0$ means $X \in \mathcal{S}_+^n$. Its dual form turns out to be:

$$\begin{aligned} \sup \quad & \sum_{i=1}^{m} b_i y_i \\ \text{s.t.} \quad & \sum_{i=1}^{m} y_i A_i + S = C \\ & S \succeq 0 \end{aligned}$$

On the modeling side, we often try to develop SDP constraint by uncovering linear combinations of (constant as well as variable) matrices in the semidefinite cone, e.g. $\sum_{i=1}^{m} \alpha_i M_i \succeq 0$ for some constants matrices M_i. Constraints in this ordered relation ($\succeq$) are referred to as Linear Matrix Inequalities, or LMIs. SDP has a relatively young history in optimization. With it computational tractability (e.g. using interior point algorithms), there is a growing interest in the field since many interesting and practical problems can be modeled as SDPs, including those in signal processing, portfolio selection problems and max-cut problems in graph theory.

5.2.3 Nonnegative polynomials with SDP

Robustness can also be discussed with polynomial constraints. Consider a polynomial $F(x) := x^n + p_1 x^{2n-1} + \cdots + p_{2n-1}x + p_{2n}$ and the statement:

$$F(x) \geq 0 \;\forall x \in \mathcal{C}, \tag{5.1}$$

where $\mathcal{C} \subseteq \mathbf{R}$ is an interval. The "for-all" condition $x \in \mathcal{C}$ is regarded as assurance of robustness. When $\mathcal{C} = \mathbf{R}$, it is well-known in the optimization literatures that (5.1) is equivalent to an LMI:

$$\begin{aligned}
& F(x) \geq 0 \;\forall x \in \mathbf{R} && (5.2)\\
\iff & \mathbf{x}^T Z \mathbf{x} \geq 0 \;\forall\; \mathbf{x} \in \mathbf{R}^{n+1} && (5.3)\\
& \text{where} \begin{cases} p_{2n} = Z_{(1,1)} \\ p_{2n-k} = \sum_{i+j=k+2} Z_{(i,j)}, \quad k = 1, \cdots, 2n-1 \\ Z_{(n+1,n+1)} = 1 \end{cases} \\
\iff & \begin{cases} Z \succeq 0 \\ p_{2n} = Z_{(1,1)} \\ p_{2n-k} = \sum_{i+j=k+2} Z_{(i,j)}, \quad k = 1, \cdots, 2n-1 \\ Z_{(n+1,n+1)} = 1 \end{cases} && (5.4)
\end{aligned}$$

In above, $Z_{(i,j)}$ denotes the ij-th element of matrix Z and the last equivalence is by the definition of semidefinite cone. When $\mathcal{C}$ is a semi-infinite or bounded interval, Nesterov [56] proved their semidefinite cone representation. Our new version of the S-Lemma in the previous chapter has further generalized the result to univariate polynomials. We quote two of the results here for the sake of convenience.

Theorem 9. *(Recall Theorem 8; see also Theorem 17.12 of Nesterov [56]) Let f and g be two univariate polynomials. Consider two statements:*

(i) $f(x) \geq 0 \qquad \forall\, g(x) \geq 0$

(ii) $\exists$ *two polynomials* $h_1(x) > 0$ *and* $h_2(x) \geq 0$, *not both constant zero, s.t.* $h_1(x)f(x) - h_2(x)g(x) \geq 0 \quad \forall\, x \in \mathbf{R}$

If g and its first derivative g' are relatively prime, then the statements are equivalent.

Lemma 8. *(Recall Corollary 1) Let $h_1(x)$ and $h_2(x)$ be defined above. Assume* $\deg g(x) \leq 2$. *Then*

$$
\begin{aligned}
\deg h_2(x) =& 0 \\
\deg h_2(x) =& 2\left(\left\lceil \frac{\deg f(x)}{2} \right\rceil - 1\right).
\end{aligned}
$$

With the results above, we are in a position to show that when $\mathcal{C}$ is a semi-infinite or bounded interval, (5.1) are LMIs as well.

Corollary 5. *When* $\mathcal{C} = (-\infty, a]$ *or* $[b, \infty)$, *(5.1) is an LMI.*

Proof. When $\mathcal{C} = (-\infty, a]$, let $g(x) := -x + a$. It is trivial that g and g' are relatively prime. By Theorem 9, there exist $h_1(x) \geq 0$ and $h_2(x) \geq 0$ such that $S(x) := h_1(x)f(x) - h_2(x)g(x) \geq 0 \ \forall x \in \mathbf{R}$. By Lemma 8, $\deg h_1(x) = 0$ and $\deg h_2(x) = 2n-2$, so we may simplify $S(x) = f(x) - h_2(x)g(x)$. Let $h_2(x) := q_0 x^{2n-2} + q_1 x^{2n-1} + \cdots + q_{2n-1}x + q_{2n-2}$. Then, similar to (5.4), (5.1) becomes:

$$F(x) \geq 0 \ \forall x \in (-\infty, a] \tag{5.5}$$

$$\iff \begin{cases} S(x) \geq 0 \ \forall x \in \mathbf{R} \\ h_2(x) \geq 0 \ \forall x \in \mathbf{R} \end{cases}$$

$$\iff \begin{cases} Z \succeq 0 \\ p_{2n} - a q_{2n-2} = Z_{(1,1)} \\ p_{2n-k} + q_{2n-k-1} + a q_{2n-k-2} = \sum_{i+j=k+2} Z_{(i,j)}, \quad k = 1, \cdots, 2n-2 \\ p_1 + q_0 = Z_{(n+1,n)} + Z_{(n,n+1)} \\ Z_{(n+1,n+1)} = 1 \\ \\ H \succeq 0, \\ q_{2n-2} = H_{(1,1)} \\ q_{2n-2-k} = \sum_{i+j=k+2} H_{(i,j)}, \quad k = 1, \cdots, 2n-3 \\ H_{(n,n)} = q_0 \end{cases} \tag{5.6}$$

This completes the proof for $\mathcal{C} = (-\infty, a]$. When $\mathcal{C} = [b, \infty)$, let $g(x) := x - b$ and repeat the arguments above. $\square$

Corollary 6. *When $\mathcal{C} = [a, b]$, (5.1) are LMIs.*

Proof. Let $g(x) := -(x-a)(x-b)$ and follow the arguments in the proof of Corollary 5. Still, we have $\deg h_1(x) = 0$ and $\deg h_2(x) = 2n-2$, but (5.1) has a different formulation of LMIs:

$$F(x) \geq 0 \; \forall x \in [a, b] \tag{5.7}$$

$$\iff \begin{cases} Z \succeq 0 \\ p_{2n} + abq_{2n-2} - t = Z_{(1,1)} \\ p_{2n-1} + abq_{2n-3} - (a+b)q_{2n-2} = Z_{(1,2)} + Z_{(2,1)} \\ p_{2n-k} + q_{2n-k} - (a+b)q_{2n-k-1} + abq_{2n-k-2} \\ = \sum_{i+j=k+2} Z_{(i,j)}, \quad k = 2, \cdots, 2n-2 \\ p_1 + q_1 - (a+b)q_0 = Z_{(n+1,n)} + Z_{(n,n+1)} \\ Z_{(n+1,n+1)} = 1 + q_0 \\ \\ H \succeq 0 \\ q_{2n-2} = H_{(1,1)} \\ q_{2n-2-k} = \sum_{i+j=k+2} H_{(i,j)}, \quad k = 1, \cdots, 2n-3 \\ H_{(n,n)} = q_0 \end{cases} \tag{5.8}$$

$\square$

An immediate application of these SDP characterization concerns the optimization of a univariate polynomial. The idea is to find a lower bound of it. It is easy to see that a real number γ is a lower bound of a polynomial $f(x)$ over $\mathcal{C}$ if and only if $f(x)-\gamma$ is nonnegative over $\mathcal{C}$. In fact,

Lemma 9. *$f(x)-\gamma$ is an SDP, given that the coefficients of $f(x)$ depend affinely on γ.*

Proof. We can find γ by formulating an optimization:

$$\begin{aligned} \max \quad & \gamma \\ \text{s.t.} \quad & f(x)-\gamma \geq 0 \quad \forall x \in \mathcal{C} \end{aligned} \tag{5.9}$$

The fact that (5.9) are LMIs is by expression (5.4), Corollaries 5 and 6 respectively. □

5.3 Some preliminaries on linear regression and confidence region

Consider a linear regression model: $y = X\beta + \epsilon$, where $y = (y_1, \cdots, y_n)^T$ is a vector of observations, $X \in \mathbf{R}^{n\times(p+1)}$ a full column-rank design matrix with the l-th $(1 \leq l \leq n)$ row given by $\mathbf{x} = (1, x_1, \cdots, x_p)$, $\beta = (\beta_0, \cdots, \beta_p)^T$ is a vector of unknown coefficients, and $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$ is a vector of independent random errors with each $\epsilon_i \sim N(0, \sigma^2)$, where σ^2 is an unknown parameter. By the least square method, we can estimate β and σ:

- $\hat{\beta} = (X^T X)^{-1} X^T y \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$ and
- $\hat{\sigma}^2 = \|y - X\hat{\beta}\|^2/(n-p-1) \sim \sigma^2 \chi^2_{n-p-1}$

Lemma 10. *(Theorem 1.2 of Liu [45]) An exact $1-\alpha$ confidence region for β is given by*

$$\left\{ \beta : \frac{(\hat{\beta}-\beta)^T (X^T X)(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2} \leq f^{\alpha}_{p+1,n-p-1} \right\}, \tag{5.10}$$

where $f^{\alpha}_{p+1,n-p-1}$ is the upper α point of the F distribution with degrees of freedom $p+1$ and $n-p-1$.

Proof. Let P be a square matrix satisfying $(X^T X)^{-1} = P^T P$, and $M = (P^T)^{-1}(\hat{\beta}-\beta)/\sigma$. Then $\mathbb{E}(M) = 0$, $Cov(M) = (P^T)^{-1} Cov(\hat{\beta}-\beta) P^{-1}/\sigma^2 = (P^T)^{-1}(X^T X)^{-1} P^{-1} = I_{p+1}$, where I_{p+1} is the $(p+1)\times(p+1)$identity matrix. So $M \sim N_{p+1}(\mathbf{0}, I)$ and $M^T M = (\hat{\beta}-\beta)^T P^{-1}(P^T)^{-1}(\hat{\beta}-\beta)/\sigma^2 = (\hat{\beta}-\beta)^T (X^T X)(\hat{\beta}-\beta)/\sigma^2$ has the chi-square distribution with $p+1$ degrees of freedom. Also note that $M^T M$ are $\hat{\sigma}^2$ are independent. So $\frac{M^T M}{(p+1)\hat{\sigma}^2} = \frac{(\hat{\beta}-\beta)^T (X^T X)(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2}$ has the F distribution with degrees of freedom $p+1$ and $n-p-1$ from which the theorem follows immediately. □

A model $\mathbf{x}^T\beta$ is plausible if and only if β is contained in the confidence region (5.10). Equivalently, a hypothesis test is designed as: $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$. H_0 is rejected if and only if $\beta_0 \notin$ (5.10). Meanwhile,

Lemma 11. *(5.10) is equivalent to the simultaneous confidence band of level $1-\alpha$ for the regression model $\mathbf{x}^T\beta$ for all $x_{(0)} := (x_1, \ldots, x_p) \in \mathbf{R}^p$, which is given by*

$$\mathbf{x}^T\beta \in \mathbf{x}^T\hat{\beta} \pm \sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}}\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}} \quad \forall\, x_{(0)} \in \mathbf{R}^p \tag{5.11}$$

Proof. Let P and M be defined in the proof of Lemma 10.

$$\begin{aligned}
&\mathbf{x}^T\beta \in \mathbf{x}^T\hat{\beta} \pm \sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}}\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}} \quad \forall\, x_{(0)} \in \mathbf{R}^p \\
\Longleftrightarrow &\sup_{x_{(0)}\in\mathbf{R}^p} \frac{|\mathbf{x}^T(\hat{\beta}-\beta|)}{\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}} \leq \sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}} \\
\Longleftrightarrow &\sup_{x_{(0)}\in\mathbf{R}^p} \frac{|(P\mathbf{x})^TM|}{(\hat{\sigma}/\sigma)\sqrt{(P\mathbf{x})^T(P\mathbf{x})}} \leq \sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}} \\
\Longleftrightarrow &\frac{||M||}{\hat{\sigma}/\sigma} \sup_{x_{(0)}\in\mathbf{R}^p} \frac{|(P\mathbf{x})^TM|}{||(P\mathbf{x})||||M||} \leq \sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}} \\
\Longleftrightarrow &\frac{||M||}{\hat{\sigma}/\sigma} \leq \sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}} \\
\Longleftrightarrow &\frac{(\hat{\beta}-\beta)^T(X^TX)(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2} \leq f^{\alpha}_{p+1,n-p-1}
\end{aligned}$$

The second to last line is due to Cauchy-Schwarz inequality. $\square$

This confidence band is the most well-known which can be found in Hoel [30] and Scheffe [64, 65]. The lower and upper parts of the band are symmetric about the fitted model

$\mathbf{x}^T\hat{\beta}$, and the width of the band at each $\mathbf{x}$ is proportional to $\sqrt{Var(\mathbf{x}^T\hat{\beta})} = \sigma\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}$, which is often referred to as hyperbolic or Scheffé type in the statistical literature. We should note that the critical value $\sqrt{(p+1)f^{\alpha}_{p+1,n-p-1}}$ is larger than the critical value t^{α}_{n-p-1} used in the pointwise confidence band for the usual hypothesis testing, as the former band has a stronger requirement ($\forall \mathbf{x}_{(0)} \in \mathbf{R}^p$). On the other hand, we still face a problem in practice with (5.11). This band requires that the linear regression model $\mathbf{x}^T\beta$ holds for all $\mathbf{x}_{(0)} \in \mathbf{R}^p$, which may be too strong a requirement when the explaining variables are only well-defined on a certain range, for instance, x_l could be ages (nonnegative) or blood pressure (in a certain range). Therefore, considering the whole range $(-\infty, \infty)$ in (5.11) may reject H_0 wrongly sometimes. This leads to a remedy by considering a more general form:

$$\mathbf{x}^T\beta \in \mathbf{x}^T\hat{\beta} \pm c\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}} \quad \forall\, \mathbf{x}_{(0)} \in \mathcal{D}, \tag{5.12}$$

where c is a suitably chosen critical constant so that the simultaneous confidence level of this band is $1-\alpha$, and $\mathcal{D}$ is a pre-specified range of concern for $\mathbf{x}_{(0)}$.

5.4 Optimization approach to the confidence bands construction

In polynomial regression, we replace $\mathbf{x}$ by $\tilde{\mathbf{x}} = (1, x_l, \cdots, x_l^p)$ in (5.12):

$$\tilde{\mathbf{x}}^T\beta \in \tilde{\mathbf{x}}^T\hat{\beta} \pm c\hat{\sigma}\sqrt{\tilde{\mathbf{x}}^T(X^TX)^{-1}\tilde{\mathbf{x}}} \quad \forall\, x \in \mathcal{C} \tag{5.13}$$

where c is a suitably chosen critical constant so that the simultaneous confidence level of this band is $1-\alpha$, and $\mathcal{C}$ is as defined above. Due to the fact that we have imposed the restriction $x_l = x^l$, the original confidence band (5.12) needs an adjustment in the sense that "$\forall\ \mathbf{x}_{(0)} \in \mathcal{D}$" is replaced by "$\forall\ x \in \mathcal{C}$", or an undesirable phenomenon may occur (see Liu et al. [47] for details).

In our optimization approach, it is crucial to note that:

Lemma 12. *Given each realization of $\xi := \frac{\hat{\beta}-\beta}{\sigma} \sim N_{p+1}(0, (X^TX)^{-1})$ and $\eta^2 := \left(\frac{\hat{\sigma}}{\sigma}\right)^2 \sim \frac{\chi^2_{n-p-1}}{n-p-1}$, (5.13) has the same representation as (5.1), namely,*

$$f_{\xi,\eta}(x) := \tilde{\mathbf{x}}^T\left[c^2\eta^2(X^TX)^{-1} - \xi\xi^T\right]\tilde{\mathbf{x}} \geq 0 \quad \forall x \in \mathcal{C} \tag{5.14}$$

Proof.

$$\begin{aligned}
&\tilde{\mathbf{x}}^T\beta \in \tilde{\mathbf{x}}^T\hat{\beta} \pm c\hat{\sigma}\sqrt{\tilde{\mathbf{x}}^T(X^TX)^{-1}\tilde{\mathbf{x}}} \\
\Longleftrightarrow &\left(\tilde{\mathbf{x}}^T(\beta - \hat{\beta}) - c\hat{\sigma}\sqrt{\tilde{\mathbf{x}}^T(X^TX)^{-1}\tilde{\mathbf{x}}}\right)\cdot \\
&\left(\tilde{\mathbf{x}}^T(\beta - \hat{\beta}) + c\hat{\sigma}\sqrt{\tilde{\mathbf{x}}^T(X^TX)^{-1}\tilde{\mathbf{x}}}\right) \le 0 \\
\Longleftrightarrow &-\left(\tilde{\mathbf{x}}^T\frac{\beta - \hat{\beta}}{\sigma}\right)^2 + c^2\left(\frac{\hat{\sigma}}{\sigma}\right)^2\tilde{\mathbf{x}}^T(X^TX)^{-1}\tilde{\mathbf{x}} \ge 0 \\
\Longleftrightarrow &f_{\xi,\eta}(x) := \tilde{\mathbf{x}}^T\left[c^2\eta^2(X^TX)^{-1} - \xi\xi^T\right]\tilde{\mathbf{x}} \ge 0,
\end{aligned}$$

where $\xi := \frac{\hat{\beta}-\beta}{\sigma} \sim N_{p+1}(0,(X^TX)^{-1})$ and $\eta^2 := \left(\frac{\hat{\sigma}}{\sigma}\right)^2 \sim \frac{\chi^2_{n-p-1}}{n-p-1}$. Hence, we arrive at (5.14). □

As an immediate application of the results in previous section,

Lemma 13. *(5.13) can be represented as LMIs if: (i) $\mathcal{C} = \mathbf{R}$; (ii) $\mathcal{C}$ is a semi-infinite interval; (iii) $\mathcal{C} = [a,b]$ (a bounded interval).*

Proof. These cases correspond to (5.4), (5.6) in Corollary 5, and (5.8) in Corollary 6 respectively. □

On the other hand, finding c in (5.13) can be posed as the following optimization problem by virtual of Lemma 12:

$$\begin{aligned}
(P) \qquad &\min \quad c^2 \\
&\text{s.t. } \mathbb{P}\{f_{\xi,\eta}(x) \ge 0 \quad \forall\, x \in \mathcal{C}\} \ge 1-\alpha, \qquad (5.15)
\end{aligned}$$

where the probability is taken with respect to both ξ and η. Since c (≥ 0) is monotone in the probability, the larger the c, the wider the confidence band, and hence larger the probability. Another remark is that minimizing c or c^2 gives the same optimal solution, given c's monotonicity and nonnegativity. Given the efficiency of computing the LMIs, we propose the following simulation scheme to find c. For the number of samples needed for Monte Carlo simulations of chance constraints, we refer to [43, 38, 67].

Simulation Scheme for c

1. Define a range $(\underline{c}, \overline{c})$ for c. Set $k = 0$ and fix N as the number of simulations for each k.

2. Set $c_k = \frac{\underline{c}+\overline{c}}{2}$.

3. Generate N pairs of independent (ξ_i, η_i^2) for $i = 1, \cdots, N$.

4. For each pair (ξ_i, η_i^2), check if

$$f_{\xi_i,\eta_i}(x) := \tilde{\mathbf{x}}^T \left[c_k^2 \eta_i^2 (X^T X)^{-1} - \xi_i \xi_i^T\right] \tilde{\mathbf{x}} \geq 0 \quad \forall x \in \mathcal{C} \tag{5.16}$$

   Set *success*=number of i that (5.16) is true.

5. Check $|\frac{success}{N} - (1-\alpha)| < \epsilon$, where $\epsilon > 0$ is predetermined. If this is true, stop and return c_k. Otherwise go to next step.

6. If $\frac{success}{N} > 1-\alpha$, then set $\overline{c} = c_k$; else $\underline{c} = c_k$. Set $k = k+1$.

7. Repeat Steps 2 to 5.

5.5 Numerical experiments

We will run two numerical experiments to test our method. The data and existing numerical results are set up in reference to Liu [45].

5.5.1 Linear regression example

This data set (page 28 of Liu [45]) is about how systolic blood pressure (Y) changes with age(x) for a group of forty males. The two extreme ages in the sample reflect the interval $[18, 70]$

of concern. Checking (5.16) refers to solving the following SDP:

$$
\begin{aligned}
(S_i) \qquad & \max \quad t_i \\
& \text{s.t.} \quad [c_k^2\eta_i^2(X^TX)^{-1} - \xi_i\xi_i^T] + h\begin{pmatrix} 18\cdot 70 & -\frac{18+70}{2} \\ -\frac{18+70}{2} & 1 \end{pmatrix} \\
& \qquad - \begin{pmatrix} t_i & 0 \\ 0 & 0 \end{pmatrix} \succeq 0 \\
& \qquad h \geq 0
\end{aligned}
$$

success = number of i that $t_i \geq 0$. Together with the results quoted in Example 2.1 of Liu [45], we showed ours in Table 5.1.

Table 5.1: 2-sided band for linear regression

reference result	2.514
Trial 1 (10000 iterations)	2.5313
Trial 2 (10000 iterations)	2.4938
Trial 3 (10000 iterations)	2.5188
Trial 4 (10000 iterations)	2.5313
Trial 5 (100000 iterations)	2.5437
Trial 6 (100000 iterations)	2.5313

5.5.2 Polynomial regression example

Another data set (page 185 of Liu [45]) is on perinatal mortality (fetal deaths plus deaths within the first month of life) rate (PMR) and birth weight (BW) collected in California in 1998. The interest is on modeling how PMR changes with BW; a 4th order polynomial regression model between $Y = \log(-\log(PMR))$ and $x = BW$ is considered. The interval of concern, based on the range of data, is $[0.85, 4.25]$.

By Lemma 8, $\deg h(x) = 2\left(\left\lceil \frac{\deg f(x)}{2} \right\rceil - 1\right) = 2 \cdot (4-1) = 6$. Let $h(x) = h_0x^6 + h_1x^5 + h_2x^4 + h_3x^3 + h_4x^2 + h_5x + h_6$. Then

$$\begin{aligned} h(x)(x-a)(x-b) =& h_0x^8 + (-(a+b)h_0 + h_1)x^7 \\ & + \sum_{k=0}^{4}(h_kab - (a+b)h_{k+1} + h_{k+2})x^{6-k} \\ & + (h_5ab - (a+b)h_6)x + h_6ab \\ =& \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \end{pmatrix}^T H \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \end{pmatrix}, \end{aligned}$$

where

$$
\begin{aligned}
h_6 ab &= H_{(1,1)} \\
h_5 ab - (a+b)h_6 &= H_{(1,2)} + H_{(2,1)} \\
(h_k ab - (a+b)h_{k+1} + h_{k+2}) &= \sum_{\substack{i+j=k+4 \\ i,j=1\cdots 5}} H_{(i,j)} \\
-(a+b)h_0 + h_1 &= H_{(5,4)} + H_{(4,5)} \\
h_0 &= H_{(5,5)}
\end{aligned}
$$

Checking (5.16) refers to solving the following SDP:

$$
\begin{aligned}
(S_i) \qquad & \max \quad t_i \\
& \text{s.t.} \quad \left[c_k^2 \eta_i^2 (X^T X)^{-1} - \xi_i \xi_i^T\right] + H \\
& \qquad - \begin{pmatrix} t_i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \succeq 0 \\
& \qquad H \succeq 0
\end{aligned}
$$

where H is defined above. The numerical results are shown in Table 5.2.

Table 5.2: 2-sided band for polynomial regression of degree 4

Reference result	2.985
Trial 1 (10000 iterations)	2.9792
Trial 2 (10000 iterations)	2.9820
Trial 3 (10000 iterations)	2.9791
Trial 4 (10000 iterations)	2.9945
Trial 5 (100000 iterations)	2.9866
Trial 6 (100000 iterations)	2.9869

5.6 Conclusion

We proposed a new computation approach to address the issue of simultaneous confidence bands in regression analysis. We focus on the 2-sided hyperbolic band in this chapter. A variety of $\mathcal{C}$ can be computed using this SDP approach, namely, $\mathcal{C}$ is the whole real, semi-infinite interval or a bounded interval. As an extension, we can also compute the hyperbolic band over a finite union of (disjoint) intervals: $\mathcal{C} = \bigcup_i^I [a_i, b_i]$, since we notice that $f(x) \geq 0 \forall x \in \mathcal{C}$ is equivalent to

$$\begin{cases} f(x) \geq 0, & \forall x \in [a_1, b_1] \\ \vdots & \\ f(x) \geq 0, & \forall x \in [a_I, b_I] \end{cases}$$

In the future, we will explore more SDP formulations for the confidence bands, in particular the 1-sided band, which is currently under investigation.

□ **End of chapter.**

Chapter 6

Moment bound of nonlinear risk measures

6.1 Introduction

6.1.1 Motivation

Without knowing the distribution of a random variable x, is it possible to estimate the expectation of the variable $\psi(x)$? As we shall see later, such problems are pervasive in risk management and financial engineering. To a lesser extent, we are interested in a confidence interval $[a, b]$ where $\mathbb{E}[\psi(x)] \in [a, b]$. In case $\mathbb{E}[\psi(x)]$ refers to a risk measure, we raise a particular concern on its (worst) upper bound. If some partial information of x is available, say its moment, then we show in this chapter that it is possible to compute such a 100% confidence interval.

While there is an opinion that such a 100% confidence interval may be too wide to be practical in view of tracking $\mathbb{E}[\psi(x)]$, we believe that this is still of the risk managers' concern, especially in extreme events. Meanwhile, if regarding $\sup \mathbb{E}[\psi(x)]$ (or $\inf \mathbb{E}[\psi(x)]$) is merely a risk measure, it is always meaningful to compare different risk measures under the same circumstances regardless of its absolute magnitude.

We believe that some partial knowledge is a practical assumption to make. As a matter of fact, extreme events are of great concern in insurance industry, as we need to worry about the risk (or the loss) in the worst situations, where an insurance coverage applies. However, owing to its low frequency of occurrence, extreme events can only be modeled or predicted with limited information. This is in contrast to the normal (or non-extreme) situations where more data or information is available for estimations and statistical analysis. Very often, if we cannot find an exact solution to the risk measures, then we try to approximate it by estimating its bounds. It is now established that there is a perfect matching between such estimation and the use of semidefinite programming (SDP) – a computationally efficient model – given that we know the moments of the risk, resulting in an efficient method to compute the moment bounds of the risk measures.

6.1.2 Robustness and moment bounds

Viewed as a conservative measure, "the worst extreme" often refers to robustness in the field of optimization. Ben-Tal et al. [2] explored rather comprehensively in the general formulation and applications in robust optimizations. They discussed when and how robust optimization methods could be formulated or approximated by SDPs. When robustness is considered in terms of probability distributions, which is known as the distributional robustness in the literature, the context turns into the generalized moment bound problem, the formulation of which is described as follows:

Let x be a random vector in $\mathbf{R}^d$. Given the moment information of $x, \mathbb{E}[x^i] = m_i, i = 0, \cdots, n$ (recall that $\mathbb{E}[x^0] = \mathbb{E}[1] \equiv 1$ is the probability of any whole sample space. Therefore $m_0 \equiv 1$ by definition), we want to find the upper (*resp.* lower) bounds on the expectation of a related quantity, $\psi(r)$, with the optimization problem:

$$
\begin{aligned}
&\text{(GP)} \quad \sup_{x\sim(m_0,\cdots,m_n)} \mathbb{E}[\psi(x)] \\
&:= \sup_{x\in\mathfrak{P}} \mathbb{E}[\psi(x)] \\
&\quad s.t. \quad \mathbb{E}[x^i] = m_i, \quad i = 0,\cdots,n \\
(\textit{resp.}) \quad &\text{(GPdown)} \quad \inf_{x\sim(m_0,\cdots,m_n)} \mathbb{E}[\psi(x)] \\
&:= \inf_{x\in\mathfrak{P}} \mathbb{E}[\psi(x)] \\
&\quad s.t. \quad \mathbb{E}[x^i] = m_i, \quad i = 0,\cdots,n
\end{aligned}
$$

where the optimization is taken over all possible distributions of the random variable x in the class $\mathfrak{P}$. It can be easily seen that the objective in (GPdown) can be regarded as $-\sup \mathbb{E}[-\psi(x)]$, which is essentially the same form as the objective in (GP). Hence the following discussion is not only applicable to the upper bounds, but also easily adapted to derive from its lower bounds. (Note that when $i = 1$, the first moment $\mathbb{E}[x]$ and m_1 are vectors; when $i = 2$, $\mathbb{E}[x^2]$ and m_2 are (covariance) matrices; when $i \geq 3$, $\mathbb{E}[x^i]$ and m_i are tenors, a higher dimension version of matrices. We will avoid discussing tenors throughout this chapter.)

Remark. *In this chapter the SDP formulations for the moment bounds problems are exact, not approximations. In general, this* *may not always be possible. Namely, not every moment bound problem will have an exact SDP formulation.*

6.1.3 Literature review in general

The feasibility of Problem (GP) given moments requirements is a classical moment problem. Well known results are available in the literature; cf. Chebyshev [9], Markov [51], Karlin and Sudden [36] and Kemperman [37]. Given the first and second moments of a random variable in $\mathbf{R}$, Chebyshev's inequality gives a bound on the distribution function. Bertsimas and Popescu [6] generalized the result by computing optimal bounds, using SDP, on arbitrary distributions given any finite number of generalized moments. He et al. [29] strengthened the inequality bounds when the first, second and fourth moments are known. Popescu [60] further generalized the SDP approach for convex classes of distributions. Zuluaga and Pena [83] worked on numerical solutions and approximations for the generalized tchebycheff inequalities. Regarding the applications of the moment bounds, Scarf [63] first derived the bound given mean-variance demand information in an inventory control problem. Lo [48] and Grundy [27] applied the result to bound the option price, where x is a stock price. In the meantime, Levy [42] discussed the option bounds with stochastic dominance approach. Rodriguez [62] developed a unified approach for several existing option bounds. Zuluaga et al. [84] extended Lo's bound to third-order. In the field of actuarial science (more details in the next subsection), Jansen [33] computed the analytical upper bound for the stop-loss payment given up to forth moment. Brockett and Cox [8] discussed the insurance calculations using incomplete information. Cox [11] developed bounds under a bounded support with a variety of piecewise linear functions. Cox et al. [12] computed and approximated the bounds of the ruin probabilities and Value-at-Risk, through the sum-of-squares (sos) formulations. Their extension can be found in Tian [73]. Schepper and Heijnen [14] derived bounds on tail probabilities and Value-at-Risk with a small number of parameters. Recently, Liu and Li [44] considered the bound by assuming a bounded support and a unimodal distribution for the univariate random variable. They also obtained a bound for the variance. Moment bounds are also applicable to the portfolio selection problems, in which $\psi(x)$ can be some utility function with x being the weight allocation of different assets (see Chen et al.[10], El Ghaoui et al.[21], Han et al.[28]). Negative components of x means short-selling the corresponding asset(s).

6.1.4 More literature review in actuarial science

Specially in the field of actuarial science, estimating moment bounds was in fact a very ad hoc issue during 1980's. At that time, a host of research papers were published, mainly by Vylder, Kaas, Goovaerts, Taylor and their co-authors (see e.g.[33, 35, 24, 18, 19, 15, 17, 16, 23, 71, 72] and the references therein). In most cases, the risk measure in question was $\psi(x) = (x - k)_+$, where k was a given constant, since this was the stop-loss expression often used in insurance and reinsurance payoffs calculations. Broadly speaking, there were two sub-streams in this line of research. While some were exclusively devoted to an analytical form of the bounds (e.g. [18, 33, 11, 19]), there were also studies focusing on numerical bounds and the underlying theories. In particular, Vylder [15] linked the moment bounds problem with convex cones, and Goovaerts et al. [23] suggested an approximation scheme using linear programming (LP). An advancement of this sub-stream led to robust optimization, which further led to semidefinite programming (SDP)– a fundamental tool underlying this chapter. Let us emphasize that the proposed formulation in this chapter is an exact solution method, not an approximation.

A serious limitation does exist in another sub-stream. It was criticized in Goovaerts et al.[23] that the method for analytical forms "can only be worked out in practice if $n \leq 3$", where n refers to the number of integral constraints (i.e. amount of given partial information). Even so, there were some recent papers on this sub-stream. Schepper and Heijnen [13] found the explicit form of bounds provided up to the first three moments and the mode. Laurence and Wang [41] derived in closed form some distribution-free bounds and optimal subreplicating strategies for spread options in a one-period static arbitrage setting. Most recently, Goovaerts et al. [25] revisited Vylder and his co-authors' results and found best-possible upper bounds on a rich class of risk measures. They also extended the discussion for the multivariate case. These papers reflect an on-going research activities on the moment bounds problem or the distribution-free computations.

6.1.5 Our contribution

Specifically, in this chapter we focus on an instance of Problem (GP) with $x \in \mathbf{R}$, where $\psi(x) = L\left(\frac{p(x)}{q(x)}\right)$ with $L(\cdot)$ being some linear or piecewise linear function, and $p(x)$ and $q(x)$ are polynomials. In words, we try to handle nonlinear risks in the form of fractional polynomials. Given the existing methodology of polynomial optimization with SDP characterizations Nesterov [56], we introduce this modern powerful tool in the framework of nonlinear risk management, or more specifically, interest rate risk management. To our knowledge, there is no previous discussion on nonlinear risk management regarding the use of the moment bounds. In subsequent sections, we will discuss the methodology of calculating the moment bound for the nonlinear quantity's expectation, followed by that of two typical risk measures associated with it, namely, the worst-case probability and the worst-case downside risk. Imagine that x is an interest rate. In a broad sense, x may represent as bond yields, mortgage rates, or any rates used in discounting future cashflows. The interest rate linked products are in the form of $L\left(\frac{p(x)}{q(x)}\right)$ including annuity life products, bond options and mortgage payments. We may not always have close form expressions for the risk or the price. Even if we do, from the risk management point of view, it is crucial to know the worst case. We will discuss how these bounds can be obtained through SDP formulations.

Given the insurance literatures in 1980's and the follow-up papers, we would like to introduce the use of SDP. Not only because it is one of the most powerful models in the field of optimization, but also the SDP formulations, like the ones in this chapter, can provide an exact numerical solution (subject to machine rounding errors) to the problems, whereas in those literatures (see Goovaerts et al. [23, 25]), only an approximation scheme was suggested.

The rest of this chapter is organized as follows. The moment bound problems, the duality theory, and some results on non-negative polynomials are introduced in section 6.2. In section 6.3, the use of SDP in moment bound applications in nonlinear form of risk is presented, under the framework of managing the interest rate risk in mortgage business, followed by some numerical results and conclusions in sections 6.4 and 6.5 respectively.

6.2 Methodological fundamentals behind the moment bounds

A well known example of moment bounds is probably the option bound provided by Lo [48]:

$$\begin{aligned}
&\sup_{x\sim(m_1,m_2)_+} \quad \mathbb{E}[x-k]_+ \\
&= \begin{cases} \mu - k\frac{m_1^2}{m_2} & \text{if } k \leq \frac{m_2}{2m_1} \\ \frac{1}{2}(m_1 - k + \sqrt{k^2 - 2m_1 k + m_2}) & \text{otherwise} \end{cases}
\end{aligned}$$

Referring to (GP), the above model takes $\psi(x) = (x-k)_+$ and assumes the knowledge of only the first two moments ($n = 2$).

The above close form solution was also known to Scarf [63], who considered almost the same function $\psi(x) = \min(x, k)$. Jansen [33] derived an explicit form given the moments information up to the fourth order, which appeared to be sophisticated. In general, given any n, a closed form expression of (GP) is only possible in rare cases. One would thus naturally be led to numerical methods. It turns out that the conic optimization approach brings fruitful results in this context. In view of this, the dual formulations and semidefinite programming (SDP) need to be introduced.

6.2.1 Dual formulations, duality and tight bounds

(GP) is an infinite dimensional problem, since the decision variable is any (discrete or continuous) probability distribution satisfying the moment requirement. This makes the problem counterintuitive to understand and is difficult to solve in general. However, if we shift our focus to its dual formulation, then the problem may yield to numerically tractable procedures in some interesting cases, depending on the dimension of x and the choice of $\psi(x)$. In this chapter, we restrict our discussion to $x \in \mathbf{R}$ and let us name the restricted (GP) as (GP1). Then the dual formulation of (GP1) is

$$
\begin{aligned}
\text{(GD1)} \qquad & \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_i m_i \\
& s.t. \quad \sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall x \in \Omega
\end{aligned}
$$

where Ω is a pre-defined sample space. It is worth noting that the dual formulation is an upper bound of its primal problem (GP'), which is guaranteed by the weak duality theorem.

Theorem 10. *(Weak Duality) Let v_d be the optimal value of (GD1), and v_p be the optimal value of (GP1). We have $v_p \leq v_d$.*

Proof. For any feasible distribution π in (GP1), and dual feasible solution $\bar{z}_0, \cdots, \bar{z}_n$ to (GD1), the dual constraint implies that

$$\mathbb{E}^{\pi}[\psi(x)] \leq \mathbb{E}^{\pi}[\sum_{i=0}^{n} \bar{z}_i x^i] = \sum_{i=0}^{n} \bar{z}_i \mathbb{E}^{\pi}[x^i] = \sum_{i=0}^{n} \bar{z}_i m_i,$$

where the first equality is given by the primal constraints. The inequality thus also holds at the respective optimalities. □

When the equity holds (i.e. $v_p = v_d$), we call v_d a tight (upper) bound. The weak duality being always true, nonetheless, the most informative upper bound is the tight one, and this is guaranteed available by the strong duality (see, e.g. Luo et al. [49], Shapiro [66] and Isii [32]).

Theorem 11. *(Strong Duality) If (GP1) is feasible and the dual (GD1) is strictly feasible, then* $v_p = v_d$*.*

Let us remark that (GP1) being feasible refers to any distribution of x satisfying the given moments while a strict feasibility of (GD1) means there exist $\bar{z}_0, \cdots, \bar{z}_n$ such that the inequalities in the constraint are strict.

6.2.2 SDP and LMIs for some dual problems

By Theorem 11, we can look into the dual problem (GD1) for the optimal solution to (GP1). The next question is whether or not (GD1) is computable. Practically, we mean to ask if the dual can be formulated by semidefinite programming (SDP[1]), or equivalently, if the dual constraint can be written as Linear Matrix Inequalities[2] (LMIs). (For more information about SDP and LMI, we refer the interested reader to Boyd and Vandenberghe [7] and Nemirovski [54]). Given that $x \in \mathbf{R}$ (univariate), the possibility of forming LMIs depends on the choice of $\psi(x)$ and Ω. When $\sum_{i=0}^{n} z_i x^i - \psi(x)$ is a polynomial over a bounded interval or semi-infinite interval, Nesterov [56] provided an affirmative

[1]Informally speaking, SDP is linear programming with LMI(s).

[2]Let $y = [y_1, \cdots, y_k]$ be a variable vector and $A_0, A_1, \cdots, A_k$ be some constant matrices. An LMI is of the form $A_0 + y_1 A_1 + \cdots + y_k A_k \succeq 0$, where "$\succeq 0$" means that the sum is a positive semidefinite matrix. Upon arriving at this form, the constraint can be handled computationally with Matlab free toolboxes, for instance, SeDuMi [69] and cvx [26].

answer by giving an explicit description of the cones of polynomials that are representable as LMIs. Subsequently, Bertsimas and Popescu [6] gave an explicit formulation for applications of moment problems. To summarize, when $\sum_{i=0}^{n} z_i x^i - \psi(x)$ is a polynomial, we conclude that each of the following three dual problems can be written as an SDP.

$$(D1) \qquad \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_i m_i$$
$$\text{s.t.} \quad \sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall x \in \mathbb{R}$$

$$(D2) \qquad \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_i m_i$$
$$\text{s.t.} \quad \sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall x \geq 0$$

$$(D3) \qquad \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_i m_i$$
$$\text{s.t.} \quad \sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall x \in [a, b]$$

Our claims are backed by Theorems 15, 16 and 17 respectively in Appendix. For the calculations of the nonlinear risk and its risk measures in this chapter, we extend our discussion to the case where the dual problem with $\psi(x)$ being a fractional polynomial and Ω a union of (disjoint) intervals. Our results stipulate that under some conditions, such models can also be written as an SDP:

Theorem 12. *Let $I_j = [a_j, b_j]$ for $j = 1, \cdots, k$, where $-\infty \leq a_1 < b_1 < a_2 < \cdots < b_k \leq \infty$. Consider*

$$(D4) \qquad \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_0 m_i$$
$$s.t. \sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall x \in \bigcup_{j=1}^{k} I_j$$

When $\psi(x)$ is a fractional polynomial, (D4) is an SDP.

Proof. Let $\psi(x) = \frac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are some polynomials and $q(x) \neq 0$. Then

$$\sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall x \in I_1 \cup \cdots \cup I_k$$
$$\iff q(x) \sum_{i=0}^{n} z_i x^i \geq p(x) \quad \forall x \in I_1 \cup \cdots \cup I_k$$
$$\iff \begin{cases} q(x) \sum_{i=0}^{n} z_i x^i \geq p(x) & \forall x \in I_1 \\ \vdots & \\ q(x) \sum_{i=0}^{n} z_i x^i \geq p(x) & \forall x \in I_k. \end{cases}$$

By Theorem 16 and 17 in the Appendix, each nonnegative univariate polynomial here can be represented with an LMI. Hence (D4) is an SDP. $\square$

Similarly, we can extend to handle the constraint of nonnegative polynomial over another:

Lemma 14. *Let $g(x)$ be a polynomial and $\psi(x)$ a fractional polynomial. Consider*

$$
\begin{aligned}
(D5) \qquad & \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_i m_i \\
& s.t. \sum_{i=0}^{n} z_i x^i \geq \psi(x) \quad \forall g(x) \geq 0
\end{aligned}
$$

(D5) is equivalent to (D4).

Proof. Note that $g(x) \geq 0 \Leftrightarrow x \in I_1 \cup \cdots \cup I_k$ for some k. Hence (D5) is an SDP. □

6.3 Worst expectation and worst risk measures on annuity payments

As our work focuses on the management of nonlinear risk of a fractional polynomial form, we try to apply the moment bounds on annuity payments, in particular on mortgages, which is a natural association of such a form where interest rate is considered a risk. With different choices of $\psi(\cdot)$, we will discuss the use of the dual forms in previous section in risk management.

6.3.1 The worst mortgage payments

Let us start with the most familiar annuity formula. Given the mortgage loan P, period t and interest rate r and the annuity A, we have

$$P = A\left(\frac{1}{1+r} + \cdots + \frac{1}{(1+r)^t}\right) = A\frac{(1+r)^t - 1}{r(1+r)^t}$$

$$f_{P,t}(r) := A = \frac{Pr(1+r)^t}{(1+r)^t - 1} \tag{6.1}$$

If we fix P and t, the annuity can be regarded as a nonlinear (fractional) polynomial in the interest rate. Before we enter into a mortgage contract, we should take the first precautious step to estimate how worst the periodic payment can be. In other words, given the moment information of r, how much is charged in the worst case? We can formulate this as the moment bound problem with $\psi(r) = f_{P,t}(r)$ as follows:

$$\begin{aligned}(RM1) \qquad & \sup_{r\sim(m_1,\cdots,m_n)_+} \mathbb{E}(f_{P,t}(r)) \\ &= \inf_{z_0,\cdots,z_n} \sum_{i=0}^{n} z_i m_i \\ &\quad \text{s.t.} \ \sum_{i=0}^{n} z_i r^i \geq f_{P,t}(r) \quad \forall r \geq 0\end{aligned}$$

By Theorem 16, (RM1) is an SDP and therefore can be computed efficiently.

6.3.2 The worst probability of repayment failure

Risk is typically viewed as an uncertainty, or a random variable to be realized, in the future. In order to get more information for what one might face in the future, people naturally seek to know the chance of each scenario's occurrence. For example, both parties, the mortgagor and the mortgagee, typically worry about the ability of the former to pay a series of periodic obligated payments. If the mortgagor can only set aside a portion of his monthly income, let say h, how likely is his failure in the mortgage obligation when the interest rate moves against him? Mathematically, he needs to estimate the greatest probability that the mortgage payment exceeds h. When we know the moments of r, this probability can be calculated by formulating a moment bound problem (i.e. an instance of Problem (GP)) as follows

$$
\begin{aligned}
(RM2) \qquad & \sup_{r \sim (m_1, \cdots, m_n)_+} \mathbb{P}(f_{P,t}(r) \geq h) \\
&= \inf_{z_0, \cdots, z_n} \sum_{i=0}^{n} z_i m_i \\
&\quad \text{s.t.} \ \sum_{i=0}^{n} z_i r^i \geq \mathbf{1}_{\{f_{P,t}(r) \geq h\}} \quad \forall r \geq 0 \\
&= \inf_{z_0, \cdots, z_n} \sum_{i=0}^{n} z_i m_i \\
&\quad \text{s.t.} \ \begin{cases} \sum_{i=0}^{n} z_i r^i \geq 1 \ \ \forall f_{P,t}(r) \geq h, r \geq 0 \\ \sum_{i=0}^{n} z_i r^i \geq 0 \ \ \forall f_{P,t}(r) < h, r \geq 0 \end{cases}
\end{aligned} \tag{6.2}
$$

Note that

$$
f_{P,t}(r) = h \Rightarrow P \cdot r(1+r)^t = h((1+r)^t - 1). \tag{6.3}
$$

Therefore, $f_{P,t}(r) \geq h$ can be represented by a union of intervals. Together with $r \geq 0$, the "for-all" ("$\forall$") condition above is still a union of intervals, thus representable as LMIs by Lemma 14. Another key to note is that $\mathbb{P}(f_{P,t}(r) \geq h) = \mathbb{E}[\mathbf{1}_{\{f_{P,t}(r) \geq h\}}]$, where $\mathbf{1}_{\{x \in \mathcal{A}\}}$ takes value 1 if $x \in \mathcal{A}$ and 0 otherwise.

6.3.3 The worst expected downside risk of exceeding the threshold

Treating the probability as a preliminary estimation on risk prevailing in the market, a financial institution (or mortgagee) needs to take more steps when accepting mortgage applications. In the subprime crisis, in order to diversify the risk of this mortgage pool, financial institutions securitized it into different tranches of bond-like hybrids, called collateralized mortgage obligations (CMOs). To make these products attractive enough, they are covered with an insurance (most of which were from AIG during the crisis) to become bonds of investment grades. As a result, most banks, pension funds and insurance firms, held with confidence a rather large portfolio of them. When the interest rate rallied, all such investment graded products became toxic assets poisoning quite a number of the entities and the disaster followed. While someone blamed the quants for facilitating the domino effect with complicated models and some blamed the human greed, we believed one of the fundamental reasons was due to the underestimation of the interest rate risk at stake. Taking into account the precautious attitudes in financial industry and the fact that it is always a difficult task to forecast the stochastic movements of interest rate, we urge for the concern about the worst expected risk of failing mortgage payments: If the payment $f_{P,t}(r)$ climbs beyond the homeowner's ability (or a certain threshold) h, what is the expected value of $[f_{P,t}(r) - h]_+$ based on the handy moment information of the interest rate? Again, we can consider the dual formulation of the moment bound problem:

$$
\begin{aligned}
(RM3) \quad & \sup_{r \sim (m_1, \cdots, m_n)_+} \mathbb{E}(f_{P,t}(r) - h)_+ \\
&= \inf_{z_0, \cdots, z_n} \sum_{i=0}^{n} z_i m_i \\
&\quad \text{s.t.} \ \sum_{i=0}^{n} z_i r^i \geq (f_{P,t}(r) - h)_+ \quad \forall r \geq 0 \\
&= \inf_{z_0, \cdots, z_n} \sum_{i=0}^{n} z_i m_i \\
&\quad \text{s.t.} \ \begin{cases} \sum_{i=0}^{n} z_i r^i \geq f_{P,t}(r) - h & \forall f_{P,t}(r) \geq h, r \geq 0 \\ \sum_{i=0}^{n} z_i r^i \geq 0 & \forall f_{P,t}(r) \leq h, r \geq 0 \end{cases}
\end{aligned}
\tag{6.4}
$$

By Lemma 14, constraints in (6.4) can be re-formulated as LMIs.

We suggest that this risk measure could be put into practice in quite some places. For instance, nowadays, mortgage applicants need to submit their credit information, such as monthly salary and loan record over the past few years. The bank in charge then makes sure that they pass a certain stress test based on their credit profile before offering the loan. The aforementioned value of (6.4) can certainly play a role in setting the passing mark there. Suppose that, under the prevailing 1M-LIBOR rate, an applicant needs to repay USD$1000 monthly, which is currently affordable based on his credit profile. Can his remaining salary and savings absorb the maximum expected high side $[f_{P,t}(r) - \$1000]_+$? If not, his application should probably be turned down or he may be required to purchase some facilities to enhance his credit.

Another application is insurance pricing, which is obvious from the very nature of its form $\mathbb{E}[f_{P,t}(r) - h]_+$. From the prospect of home buyers, floating rate mortgage plan is more attractive than the fixed rate plans in the low interest environment. But what if the interest rate soars? Although the mortgage plan usually includes a cap on the rate, this still possibly creates a serious financial burden for them. The reason is that this cap is usually referenced to another more stable and yet changing rate (e.g. the PRIME rate in Hong Kong). Meanwhile, the cap may still be too high for protection. Being offered an insurance against the unwanted high side of the pre-specified amount, the home buyer can decide to pay the extra premium for the protection. In other words, we mean to calculate $\mathbb{E}[\sum_{j=1}^{J} \delta_j [f_{P,t}(r) - h]_+]$, where δ_j is some discount factor for the j-th repayment. Since $f_{P,t}(r)$ is nonlinear, there is no easy way to find its close form solution, but the upper bound can be estimated by our method with some common choices of δ_j:

1. If the discount factor δ_j is chosen to be independent of r, we can simply compute
$$\sum_{j=1}^{J} \delta_j \sup_{r\sim(m_1,\cdots,m_n)_+} \mathbb{E}[[f_{P,t-j+1}(r) - h]_+].$$

2. If the discount factor is $\delta_j = \frac{1}{(1+r)^j}$, we will have to handle some of piecewise fractional polynomials $\sum_{j=1}^{J} \frac{1}{(1+r)^j}[f_{P,t-j+1}(r) - h]_+$ as our objective. This is sophisticated, but still computable with our models.

3. If the discount factor is $\delta_j = \frac{1}{(1+r+s)^j}$, i.e. depending on r plus a given constant spread s, it is easy to see from 2 that our model still applies.

Therefore the numerical value of the bounds could be obtained.

6.4 Numerical examples for risk management

6.4.1 A mortgage example

To demonstrate the use of our models, let us present an experimental scenario.

Consider an annuity for a loan \$1000 with annual payments in the next 20 years, charged for a floating interest rate. Suppose the latest reference rate is 2.5% p.a.. By (6.1), the annual payment

$$f_{1000,20}(0.0013) = \$50.69.$$

To mimic the trend of interest rate in real situation, we took the mean and standard deviation based on the historical 1-year Treasury rate[3]. In particular, we chose the most recent 5-year (2007-2011), 10-year (2002-2011) and 20-year (1992-2011) monthly samplings for comparisons, which reflected concerns on different risk horizons. According to our first model, each of the three sets of data gives a tight bound on $\sup \mathbb{E}[f_{1000,20}(r)]$: If

Table 6.1: Worst expectation in different periods.

Sampling period	μ	σ	$\sup \mathbb{E}[f_{1000,20}(r)]$	$\frac{\sup \mathbb{E}[f_{1000,20}(r)]}{f_{1000,20}(0.025)} - 1$
5-year	1.46%	1.70%	\$58.4817	15%
10-year	2.10%	1.68%	\$62.1876	23%
20-year	3.52%	2.00%	\$71.1213	40%

this loan serves a period of 5 and 10 years, our model shows that the risk (potential increase) can be as much as 15% and 23% of the current level respectively. In case the loan is taken until its

[3] 1-year Treasury constant maturities is used and can be obtained from `http://www.federalreserve.gov/releases/h15/data.htm`

maturity, our model based on a 20-year sampling suggests that there is a risk of increased payment up to 40% of the current level.

Meanwhile, one of the common risk management techniques is to study the risk at one or two standard deviation about the mean. We can convert these risk levels into corresponding thresholds h. In other words, we worry about the increased payments due to interest rate fluctuations. When the payment reaches a certain threshold level h, we may need coverage or cede the unwanted risk to other parties. Based on our second and third models, we can calculate the maximum stop loss payment as well as probability with our models at the thresholds.

Table 6.2: Risk measures in different periods and different thresholds.

Sampling period	$\mu+\sigma$	eqv. threshold h^4	$\sup \mathbb{E}[f_{1000,20}(r)-h]_+$	$\sup \mathbb{P}(f_{1000,20}(r)\geq h)$
5-year	3.16%	\$61.2121	\$2.3312	0.4630
10-year	3.78%	\$72.1465	\$2.3666	0.5000
20-year	5.53%	\$83.8605	\$3.0463	0.5000
Sampling period	$\mu+2\sigma$	eqv. threshold h^5	$\sup \mathbb{E}[f_{1000,20}(r)-h]_+$	$\sup \mathbb{P}(f_{1000,20}(r)\geq h)$
5-year	4.86%	\$79.2686	\$1.4726	0.2000
10-year	5.46%	\$83.3836	\$1.4767	0.2000
20-year	7.53%	\$98.3171	\$1.8929	0.2000

Compare the results of the two risk levels $\mu+\sigma$ and $\mu+2\sigma$. If we manage to retain more risk, the coverage can be as cheap as

[4] $h = f_{1000,20}(\mu+\sigma)$

[5] $h = f_{1000,20}(\mu+2\sigma)$

63% of otherwise (e.g. $\frac{1.4726}{2.3312} \approx 63\%$). If we consider the worst probabilities as the payoff of binary options $\mathbb{E}[\mathbf{1}_{f_{1000,20}(r)\geq h}]$, a similar conclusion follows (40% in this case). Equipped with the calculated chance of events and the cost for protections, we may judge the acceptance level of risk.

6.4.2 An annuity example

When pricing a policy, actuaries need to assume the interest rate charged. We can go through a sensitivity analysis with the following simplified example. Suppose an actuary wants to price, for a male of age 50, an annuity life insurance which will offer ten annual payments of $5000. If he passes away during the period, the annuity terminates immediately, but a death benefit of $50000 will be provided.

Let us refer to the 2007 period life table in the U.S. Social Security Administration.[6] The actuarial present value for this policy is calculated to be $v = \$45259.40$ (calculated in the following table) assuming a 5% annual interest rate. We further assume that there is no loading. Then v is the required reserve which is set against the man's mortality in the insurance company. To justify the interest rate of 5%, the actuary can use the moment

[6] `http://www.ssa.gov/oact/STATS/table4c6.html`

Table 6.3: Calculations of the actuarial present value of the policy.

policy year i	survival rate at i	mortality rate at i (q_i)	$\alpha_i := 5000p_i + 50000q_i$	$\frac{1}{1.05^i}$	discounted paid off
0	0.99449	0.00551	5248.04	1.0000	5284.04
1	0.98851	0.00597	5241.31	0.9524	4991.72
2	0.98209	0.00643	5231.71	0.9070	4745.31
3	0.97524	0.00685	5218.78	0.8638	4508.18
4	0.96796	0.00727	5203.39	0.8227	4280.84
5	0.96024	0.00772	5187.20	0.7835	4064.31
6	0.95203	0.00821	5170.72	0.7462	3858.47
7	0.94330	0.00873	5153.20	0.7107	3662.28
8	0.93401	0.00929	5134.53	0.6768	3475.25
9	0.92412	0.00989	5114.90	0.6446	3297.11
10	0.91358	0.01054	5094.99	0.6139	3127.88
				Premium	$45259.40

bound to estimate the worst expectation[7] with a reference rate, say, the 1-year market yield on U.S. Treasury[8] securities. In this case, (RM1) can be applied by considering $\tilde{f}(r) := \sum_{i=0}^{10} \frac{\alpha_i}{(1+r)^i}$ and with the first two moments, μ and $\mu^2 + \sigma^2$. From the data set, we further obtain the range: $r \in [0.18\%, 14.18\%]$. A transformation $x = 1 + r$ (so that we use $f(x) := \tilde{f}(x-1) = \sum_{i=0}^{10} \frac{\alpha_i}{x^i}$) will lead to this moment bound formulation:

[7]Here assumes that the randomness of interest rate is independent of mortality.

[8]`http://www.federalreserve.gov/releases/h15/data.htm`

$$
\begin{aligned}
&\sup_{x\sim(1+\mu,(1+\mu)^2+\sigma^2)_+} \mathbb{E}(f(x)) \\
&= \inf_{z_0,z_1,z_2} z_0 + z_1(1+\mu) + z_2((1+\mu)^2+\sigma^2) \\
&\quad \text{s.t. } z_0 + z_1 x + z_2 x^2 \geq f(x) \quad \forall x \in [1.0018, 1.1418]
\end{aligned}
$$

The actuary can compare the impact of interest rate according to different horizons of history on the reserve. The calculation is summarized in the following table: In the table, the

Table 6.4: Worst case risk in different periods.

Horizon of History	μ	σ	$\sup \mathbb{E}[f(x)]$	% of risk
5 years	1.466%	1.836%	\$53396.00	18%
10 years	2.102%	1.721%	\$51789.40	14%
20 years	3.524%	2.000%	\$48621.50	7%

risk is regarded as the portion in excess of the reserve. The risk using 20 year's history is lower. This may be explained by the fact that the 20-year-set reflects the higher possibility of high interest rate environment, so that it is easier to generate more interest return to meet the annuity obligations.

6.5 Conclusion

By Theorem 12, we manage to extend the SDP techniques to calculate moment bounds for nonlinear risk in the form of fractional polynomials. When we estimate a variety of risk measures or payments in such nonlinear forms, we can use the techniques to find their supremum (and infimum, if necessary). We believe that these estimations are essential in risk management, especially when we have limited information on the products.

□ **End of chapter.**

Chapter 7

Computing distributional robust probability functions

In risk management, we are concerned about the worst-case analysis. With reference to a risk measure, one of the most popular ways to describe "worst-case" is through distributional robustness, which refers to any possible distribution from a given set. The given set is often described by some moment information, say $\mathfrak{m}_0, \mathfrak{m}_1, \cdots, \mathfrak{m}_n$, and the formulation as follows is referred as *moment bound problem*:

$$
\begin{aligned}
(GP) \quad & \sup_{x \sim (\mathfrak{m}_0, \cdots, \mathfrak{m}_n)} \mathbb{E}[\psi(x)] \\
& := \sup \quad \mathbb{E}[\psi(x)] \\
& \quad\quad s.t. \quad \mathbb{E}[\underbrace{x \odot x \odot \cdots \odot x}_{\#\text{ of } x = i}] = \mathfrak{m}_i, \quad i = 0, \cdots, n
\end{aligned}
$$

(When $x \in \mathbf{R}$, "$\odot$" is the scalar multiplication and $\mathfrak{m}_i \in \mathbf{R}$ for all i, we denote $\mathbb{E}[x^i] = \mathfrak{m}_i$; When $x \in \mathbf{R}^d$, "$\odot$" is the tensor multiplication (or matrix multiplication when $d = 2$) in the corresponding spaces and $\mathfrak{m}_i \in \mathbf{R}^{\overbrace{d \times \cdots \times d}^{\#\text{ of } d=i}}$. For example, if $x = [x^{(1)}, x^{(2)}]^T \in \mathbf{R}^2$, then $x \odot x = xx^T = \begin{pmatrix} (x^{(1)})^2 & x^{(1)}x^{(2)} \\ x^{(1)}x^{(2)} & (x^{(2)})^2 \end{pmatrix} \in \mathbf{R}^{2\times 2}$, which is in the same space as $\mathfrak{m}_2$. We will use the small letters (e.g. m) for scalars and vectors, capital letters (e.g. M) for matrices, and *fraktur* small letters $\mathfrak{m}$ for ambiguous (implied) dimensions)

Scarf [63] first applied this worst-case analysis in inventory management, where he took $\psi(x) = \min\{x, k\}$ for some constant k and assumed the knowledge of the first two moments. Lo [48] and Grundy [27] applied the similar concept for option bounds. As a matter of fact, a sizeable amount of relevant literature can be found (see e.g. Chen et al. [10], Cox [11], Cox et al. [12], Han et al. [28], He et al. [29], Jansen [33], Liu and Li [44], Schepper and Heijnen [13, 14], Vylder and Goovaerts [18, 19] and Vylder [15]). With the recent computational developments of moment bounds, applications have been introduced in different streams in financial engineering. For example, Bertsimas and Popescu [6] linked the moment bounds with semidefinite programming and probability theory; Popescu [61] worked out the mean-covariance solutions for stochastic optimization; Chen et al. [10] and Natarajan and Sim [53] discussed the portfolio selection; Wong and Zhang [76] introduced the context into nonlinear risk management; Lasserre et al. [40] priced a class of exotic options with moments and SDP relaxation. Regarding the theories of computing moment bounds, readers may refer to Popescu [60] and Lasserre [39].

In particular, when we choose $\psi(x) = \mathbf{1}_{x\in\mathfrak{E}}$ for some event $\mathfrak{E}$ in the sample space $\Omega \subseteq \mathbf{R}^d$ of x, (GP) is the worst-case probability, which can be regarded as an implicit function of the moments, given the event $\mathfrak{E}$.

$$\mathbb{F}_{d,n}(\mathfrak{E}) := \sup_{(\mathbf{R}^d \ni) x \sim (\mathfrak{m}_0, \cdots, \mathfrak{m}_n)} \mathbb{P}[x \in \mathfrak{E}] \quad \left(= \sup_{x \sim (\mathfrak{m}_0, \cdots, \mathfrak{m}_n)} \mathbb{E}[\mathbf{1}_{x\in\mathfrak{E}}] \right) \tag{7.1}$$

We will show that, by choosing $\mathfrak{E} = \{x \in \mathbf{R} : x \le t\}$, $\mathbb{F}_{1,2}(\mathfrak{E})$ is a probability distribution, which may not be true in general. Although we always have $0 \le \mathbb{F}_{d,n} \le 1$, it may not satisfy the additivity of joint countable union, namely, for any countable sequence of pairwise disjoint event $\mathfrak{E}_1, \mathfrak{E}_2, \cdots$, we only have

$$\mathbb{F}_{d,n}\left(\bigcup_{j}^{\infty} \mathfrak{E}_j\right) \le \sum_{j=1}^{\infty} \mathbb{F}_{d,n}(\mathfrak{E}_j)$$

In other words, only subadditivity is guaranteed and the equality holds only when the right hand side is attained by the same extremal distribution of x for all $\mathfrak{C}_j$.

The possibility of both the analytical form and computation of $\mathbb{F}_{d,n}$ remains open for a general n and d. Throughout this chapter, we fix $n = 2$ unless specified otherwise. When $d = 1$ and $n = 2$, there are nice distributional robust functions in analytical form. We will revisit them with discussing the Value-at-Risk in the context of portfolio selection. The formulation is in line with El Ghaoui et al. [21], who discuss the worst-case Value-at-Risk with unknown first two moments. When $d = 2$, there are no analytical form or method of exact computation to our best of knowledge. The closest approximation is due to Cox et al. [12], who use sum-of-squares (sos) polynomials to approximate $\mathbb{F}_{2,2}(\mathfrak{C})$ for nonnegative random variables, where $\mathfrak{C} = \{x \in \mathbf{R}^2 : x \leq t \text{ for some } t \in \mathbf{R}^2_+\}$. Our key contribution is to provide the exact computational methods, in the form of SDP, for $\mathbb{F}_{2,2}(\cdot)$. The methodology is based on the nice characterization of copositive cones in $\mathbf{R}^{d+1}$, where $d \leq 3$, and some results in Luo et al. [50], which states that, given either (i) $x^{(1)} \in [0, 1]$ or (ii) $x^{(1)} \in \mathbf{R}_+$, and $x^{(2)} \in \mathbf{R}^m$, a bi-quadratic function can be checked the nonnegativity with LMIs.

The rest of this chapter is organized as follows. In section 7.1, we review on $\mathbb{F}_{1,2}$ and its Value-at-Risk in the context of portfolio selection. In section 7.2, we derive the LMIs for computing $\mathbb{F}_{2,2}$, where three events are taken into account as our "base cases": $\mathfrak{E}_1 := \{x \in \mathbf{R}^2 : x^{(1)} \leq u^{(1)}, x^{(2)} \leq u^{(2)}\}, \mathfrak{E}_2 := \{x \in \mathbf{R}^2 : l^{(1)} \leq x \leq u^{(1)}, l^{(2)} \leq x \leq u^{(2)}\}$ and $\mathfrak{E}_3 := \{x \in \mathbf{R}^2 : x^{(1)} \leq u^{(1)}, 1_2 \leq x \leq u^{(2)}\}$. Model extensions are introduced in section 7.3, followed by applications in section 7.4. A conclusion is in section 7.5.

7.1 Distributional robust function with a single random variable

Take $\mathfrak{E}_1 = \{x \in \mathbf{R} : x \leq t\}$ and let μ_1 and σ^2 be the mean and variance respectively. $\mathbb{F}_{1,2}(\mathfrak{E}_1)$ can be represented as a function of t (see Chen et al. [10]):

$$\mathbb{F}_{1,2}(\mathfrak{E}_1) := F(t) = \begin{cases} \frac{\sigma^2}{(\mu_1 - t)^2 + \sigma^2}, & t \leq \mu_1; \\ 1, & t > \mu_1. \end{cases} \tag{7.2}$$

This essentially comes from Chebyshev-Cantelli inequality. It is also well-known that this worse-case probability is achieved by a two-point distribution of x. However, the story is completely different when $F(t)$ is regarded as a distribution function of some random variable ζ, since it now has a smooth and continuous distribution (7.2), which allows us to compute its moments. It is interesting to note that the first two moments of ζ and x are no longer the same: $\mathbb{E}(\zeta) = \mu_1 - \frac{\pi}{2}\sigma$ versus $\mathbb{E}(x) = \mu_1$; and $\mathbb{E}(\zeta^2) = \infty$ versus $\mathbb{E}(x^2) = \mu_1^2 + \sigma^2$ (see Appendix). This infinite variance provides us with some insight about the huge "fluctuation" of ζ.

In risk management, as extreme events associate with the Value-at-Risk, let us apply ζ with this risk measure and consider a portfolio selection problem. Suppose that $\theta \in \mathbf{R}^p$ be the vector of investment return from p assets with a mean $m \in \mathbf{R}^p$ and second moment matrix $M \in \mathcal{S}_+^p$. Let $w \in \mathbf{R}^p$ be the portfolio weights and $x = w^T\theta$ the portfolio return. Then $\mathbb{E}(x) = w^T m$ and $\mathbb{E}(x^2) = w^T M w$. Applying $\mathbb{F}_{1,2}(\mathfrak{C}_1)$ in the definition VaR, where we regard $-w^T\theta$ as the loss and choose $t = -\alpha$ in $\mathfrak{C}_1$, we have

$$VaR_\epsilon(w^T\theta) := \arg\min_\alpha\{\mathbb{F}_{1,2}(-w^T\theta \geq \alpha) \leq \epsilon\},$$

where $\epsilon \in (0,1)$ is the level of confidence. The higher the α, the higher the risk. Therefore we would like to minimize the risk over the set of admissible portfolio $\mathcal{W}$ (which typically incorporates the target of return, budget constraint and sometimes no short selling) as follows:

$$
\begin{aligned}
\min \quad & \alpha \\
s.t. \quad & \left(\sup_{x \sim (w^T m, w^T M w)} \mathbb{P}(x \geq -\alpha) = \right) F(-\alpha) \leq \epsilon \qquad (7.3) \\
& w \in \mathcal{W},
\end{aligned}
$$

where ϵ is given. We are going to show that the risk constraint (7.3) is convex and efficiently computable in the following Lemma.

Lemma 15. *(7.3) is in a second-order cone (SOC).*

Proof.

$$
\begin{aligned}
F(-\alpha) &\leq \epsilon \\
\frac{w^T M w}{(w^T m + \alpha)^2 + w^T M w} &\leq \epsilon \\
(1-\epsilon) w^T M w &\leq \epsilon (w^T m + \alpha)^2 \\
\begin{pmatrix} w^T m + \alpha \\ \frac{1-\epsilon}{\epsilon} M^{\frac{1}{2}} w \end{pmatrix} &\in SOC(d+1),
\end{aligned}
$$

where $M^{\frac{1}{2}} M^{\frac{1}{2}} = M$. □

Note that we have implicitly assumed $-w^T m \leq \alpha$. Otherwise $\mathbb{F}_{1,2}(-w^T \theta \geq \alpha) = 1 > \epsilon$, which contradicts the definition of VaR. Our result is in line with that in El Ghaoui et al. [21], who arrived at the same conclusion from a completely different angle.

For completeness, let us state the worst-case probability of the event $\mathfrak{E}_2 := \{x \in \mathbf{R} : l \leq x \leq u\}$:

$$\mathbb{F}_{1,2}(\mathfrak{E}_2) = \begin{cases} \frac{\sigma^2}{(l-\mu_1)^2+\sigma^2}, & \mu_1 < l; \\ 1, & l \leq \mu_1 \leq u; \\ \frac{\sigma^2}{(\mu_1-u)^2+\sigma^2}, & \mu_1 > u. \end{cases} \tag{7.4}$$

7.2 Moment bound of joint probability

In this section, we consider three "base-case" joint events:

$$\begin{aligned} \mathfrak{E}_1 :=& \{x \in \mathbf{R}^2 : x^{(1)} \leq u^{(1)}, x^{(2)} \leq u^{(2)}\} \\ \mathfrak{E}_2 :=& \{x \in \mathbf{R}^2 : l^{(1)} \leq x^{(1)} \leq u^{(1)}, l^{(2)} \leq x^{(2)} \leq u^{(2)}\} \\ \mathfrak{E}_3 :=& \{x \in \mathbf{R}^2 : x^{(1)} \leq u^{(1)}, l^{(2)} \leq x^{(2)} \leq u^{(2)}\} \end{aligned}$$

Our goal is to show the LMI formulations for $\mathbb{F}_{2,2}(\mathfrak{E}_k)$, $k = 1, 2, 3$. Let $l := (l^{(1)}, l^{(2)})^T$ and $u := (u^{(1)}, u^{(2)})^T$ and the mean $\mu \in \mathbf{R}^2$ and covariance matrix $\Gamma \in \mathcal{S}^2_+$ of x be given. Recall their primal form,

$$
\begin{aligned}
(P_k) \qquad \sup_{x\sim(\mu,\Gamma)} \mathbb{P}[x \in \mathfrak{E}_k] := \sup_{x\in\mathfrak{P}} & \mathbb{E}[\mathbf{1}_{\mathfrak{E}_k}] \\
s.t. \quad & \mathbb{E}[x] = \mu \\
& \mathbb{E}[xx^T] = \Gamma + \mu\mu^T
\end{aligned}
$$

Since we can pick any feasible distribution from $\mathfrak{P}$ for the optimal, the bound is known as distributional robust. Another remark is that (P_k) is an infinite dimensional problem that is not trivial to solve. Therefore, we will look into their dual formulation:

$$
\begin{aligned}
(D_1) \qquad & \inf_{z_0,z_1,Z_2} z_0 + z_1^T\mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall x \leq u \qquad (7.5) \\
& \qquad\;\; z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x \in \mathbf{R}^2 \\
(D_2) \qquad & \inf_{z_0,z_1,Z_2} z_0 + z_1^T\mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall l \leq x \leq u \qquad (7.6) \\
& \qquad\;\; z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x \in \mathbf{R}^2
\end{aligned}
$$

$$
\begin{aligned}
(D_3) \qquad & \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall x^{(1)} \leq u^{(1)}, l^{(2)} \leq x^{(2)} \leq u^{(2)} \\
& \qquad\quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x \in \mathbf{R}^2
\end{aligned} \tag{7.7}
$$

Trivially, their second constraint are the same and is an LMI. In the mean time, (7.5) is a copositive constraint in dimension 3 × 3, thus also an LMI. We will supplement the derivation for completeness; To show that (7.6) and (7.7) can be cast into LMIs as well, we base the results on Theorems 13 and 14:

Theorem 13. *(Theorem 4.5 of Luo et al. [50]) Let $p(x, y) := y^T C y + 2(y^T B y)x + (y^T A y)x^2$ be defined by $p : \mathbf{R}_+ \times \mathbf{R}^m \mapsto \mathbf{R}$, where A, B, C are sub-matrices in $Z \in \mathcal{L}_{2,m}$ and*

$$
\mathcal{L}_{2,m} := \left\{ \begin{pmatrix} C & B \\ B & A \end{pmatrix} \in \mathcal{S}^{2\times m} : A, B, C \in \mathcal{S}^m \right\}.
$$

Then

$$\begin{aligned}
& p(x,y) \geq 0 \quad \forall\, x \in \mathbf{R}_+, y \in \mathbf{R}^m \\
\Longleftrightarrow Z \in & \left\{ \begin{pmatrix} C & B \\ B & A \end{pmatrix} \in \mathcal{L}_{2,m} : \right. \\
& \begin{pmatrix} C & B \\ B & A \end{pmatrix} - \begin{pmatrix} 0 & E \\ E^T & 0 \end{pmatrix} \succeq 0, \\
& \left. E + E^T \succeq 0 \text{ for some } E \right\}
\end{aligned}$$

Theorem 14. *(Theorem 4.6 of Luo et al. [50]) Let* $p(x,y) := y^T C y + 2(y^T B y)x + (y^T A y)x^2$ *be defined by* $p : [0,1] \times \mathbf{R}^m \mapsto \mathbf{R}$*, where* A, B, C *are sub-matrices in* $Z \in \mathcal{L}_{2,m}$*. Then*

$$\begin{aligned}
& p(x,y) \geq 0 \quad \forall\, x \in [0,1], y \in \mathbf{R}^m \\
\Longleftrightarrow Z \in & \left\{ \begin{pmatrix} C & B \\ B & A \end{pmatrix} \in \mathcal{L}_{2,m} : \right. \\
& \begin{pmatrix} C & B - E \\ B - E^T & A + E + E^T \end{pmatrix} \succeq 0, \\
& \left. E + E^T \succeq 0 \text{ for some } E \right\}
\end{aligned}$$

The key to apply Theorems 13 and 14 into the moment bounds is that we choose $y = (1, \xi, \cdots, \xi^m)^T$, where $\xi \in \mathbf{R}$, so that $p(x,\xi)$ has a degree $2m-2$ in ξ. In other words, it is no longer bi-quadratic. We are going to show the "conversion" and how the theorems are invoked.

7.2.1 Constraint (7.5) in LMIs

Let $x = u - \tilde{x}$. Then rewrite (7.5) with a few lines of algebra:

$$
\begin{aligned}
& z_0 - 1 + z_1^T(u - \tilde{x}) + (u - \tilde{x})^T Z_2 (u - \tilde{x}) \\
=& z_0 - 1 + z_1^T u + u^T Z_2 u - (z_1^T + 2u^T Z_2)\tilde{x} + \tilde{x}^T Z_2 \tilde{x} \geq 0 \quad \forall \tilde{x} \in \mathbf{R}_+^2 \\
\Longleftrightarrow & \begin{pmatrix} z_0 - 1 + z_1^T u + u^T Z_2 u & -z_1^T/2 - u^T Z_2 \\ -z_1/2 - Z_2 u & Z_2 \end{pmatrix} - N \in \mathcal{S}_+^3,
\end{aligned}
$$

where $N \in \mathbf{R}_+^{3\times 3}$. Here we use the fact that the copositive cone $\mathcal{C}^m = \mathcal{S}_+^m + \mathbf{R}_+^{m\times m}$ for $m \leq 4$. Hence (D_1) can be cast as an SDP:

$$
\begin{aligned}
(SDP_1) \quad & \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
s.t. & \begin{cases} \begin{pmatrix} z_0 - 1 + z_1^T u + u^T Z_2 u & -z_1^T/2 - u^T Z_2 \\ -z_1/2 - Z_2 u & Z_2 \end{pmatrix} - N \succeq 0, \\ N^{(ij)} \geq 0, \quad i, j = 1, 2, 3 \\ \begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} \succeq 0 \end{cases}
\end{aligned}
$$

Let us remark that (D_1) can be extended to compute $\mathbb{F}_{3,2}(\mathfrak{C}_1)$.

7.2.2 Constraint (7.6) in LMIs

Let $x^{(1)} = (u^{(1)} - l^{(1)})\eta + l^{(1)}$ and $x^{(2)} = \frac{l^{(2)} + u^{(2)}\xi^2}{1+\xi^2}$. Writing the expression in (7.6) componentwise and multiplying $(1 + \xi^2)^2$ on it, we have

$$
\begin{aligned}
&(1+\xi^2)^2(z_0 - 1 + z_1^T x + Z_2 \bullet xx^T) \\
=&(1+\xi^2)^2(z_0 - 1 + z_1^{(1)}x^{(1)} + z_1^{(2)}x^{(2)} + Z_2^{(11)}(x^{(1)})^2 \\
&+ (Z_2^{(12)} + Z_2^{(21)})x^{(1)}x^{(2)} + Z_2^{(22)}(x^{(2)})^2) \\
=&\Big(z_0 - 1 + z_1^{(1)}((u^{(1)} - l^{(1)})\eta + l^{(1)}) \\
&+Z_2^{(11)}((u^{(1)} - l^{(1)})\eta + l^{(1)})^2\Big)(1 + 2\xi^2 + \xi^4) \\
&+ \left[z_1^{(2)} + \left(Z_2^{(12)} + Z_2^{(21)}\right)((u^{(1)} - l^{(1)})\eta + l^{(1)})\right] \cdot \\
&(l^{(2)} + u^{(2)}\xi^2)(1+\xi^2) + Z_2^{(22)}(l^{(2)} + u^{(2)}\xi^2)^2 \\
=&c_1(\xi) + b_1(\xi)(u^{(1)} - l^{(1)})\eta + a_1(\xi)(u^{(1)} - l^{(1)})^2\eta^2,
\end{aligned}
$$

where

$$
\begin{aligned}
c_1(\xi) :=& \left(z_0 - 1 + z_1^{(1)}l^{(1)} + Z_2^{(11)}(l^{(1)})^2\right)(1+\xi^2)^2 \\
&+ \left[z_1^{(2)} + l^{(1)}\left(Z_2^{(12)} + Z_2^{(21)}\right)\right](l^{(2)} + u^{(2)}\xi^2)(1+\xi^2) \\
&+ Z_2^{(22)}(l^{(2)} + u^{(2)}\xi^2)^2 = \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}^T C_1 \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
C_1 \in &\mathfrak{C}(l^{(1)}, l^{(2)}, u^{(2)}) := \\
&\Big\{ Y \in \mathcal{S}^3 \mid Y^{(11)} = \left(z_0 - 1 + z_1^{(1)} l^{(1)} + Z_2^{(11)} (l^{(1)})^2 \right) \\
&+ \left[z_1^{(2)} + l^{(1)} \left(Z_2^{(12)} + Z_2^{(21)} \right) \right] l^{(2)} + Z_2^{(22)} (l^{(2)})^2; \\
&Y^{(13)} + Y^{(22)} + Y^{(31)} = 2 \left(z_0 - 1 + z_1^{(1)} l^{(1)} + Z_2^{(11)} (l^{(1)})^2 \right) \\
&+ \left[z_1^{(2)} + l^{(1)} \left(Z_2^{(12)} + Z_2^{(21)} \right) \right] (l^{(2)} + u^{(2)}) \\
&+ 2 Z_2^{(22)} l^{(2)} u^{(2)}; \\
&Y^{(33)} = \left(z_0 - 1 + z_1^{(1)} l^{(1)} + Z_2^{(11)} (l^{(1)})^2 \right) \\
&+ \left[z_1^{(2)} + l^{(1)} \left(Z_2^{(12)} + Z_2^{(21)} \right) \right] u^{(2)} + Z_2^{(22)} (u^{(2)})^2; \\
&Y^{(12)} = Y^{(21)} = Y^{(23)} = Y^{(32)} = 0 \Big\},
\end{aligned}
$$

$$
\begin{aligned}
b_1(\xi) := &\left(z_1^{(1)} + 2 Z_2^{(11)} l^{(1)} \right) (1 + \xi^2)^2 \\
&+ \left(Z_2^{(12)} + Z_2^{(21)} \right) (l^{(2)} + u^{(2)} \xi^2)(1 + \xi^2) \\
= &\begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}^T B_1 \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
B_1 \in &\mathfrak{B}(l^{(1)}, l^{(2)}, u^{(2)}) := \\
&\Big\{ Y \in \mathcal{S}^3 \mid Y^{(11)} = \left(z_1^{(1)} + 2 Z_2^{(11)} l^{(1)} \right) + \left(Z_2^{(12)} + Z_2^{(21)} \right) l^{(2)}; \\
&Y^{(13)} + Y^{(22)} + Y^{(31)} = 2 \left(z_1^{(1)} + 2 Z_2^{(11)} l^{(1)} \right) + \left(Z_2^{12} + Z_2^{21} \right) (l^{(2)} + u^{(2)}); \\
&Y^{(33)} = \left(z_1^{(1)} + 2 Z_2^{(11)} l^{(1)} \right) + \left(Z_2^{(12)} + Z_2^{(21)} \right) u^{(2)}; \\
&Y^{(12)} = Y^{(21)} = Y^{(23)} = Y^{(32)} = 0 \Big\},
\end{aligned}
$$

$$a_1(\xi) := Z_2^{(11)}(1+\xi^2)^2 = \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}^T A_1 \begin{pmatrix} 1 \\ \xi \\ xi^2 \end{pmatrix},$$

$$\begin{aligned} A_1 \in \mathfrak{A} := \{ Y \in \mathcal{S}^3 \mid Y^{(11)} = Z_2^{(11)}; Y^{(13)} + Y^{(22)} + Y^{(31)} = 2Z_2^{(11)}; \\ Y^{(33)} = Z_2^{(11)}; Y^{(12)} = Y^{(21)} = Y^{(23)} = Y^{(32)} = 0\}. \end{aligned}$$

By Theorem 14, constraint (7.6) can be represented by LMIs:

$$\begin{aligned} & z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall\, 0 \leq x \leq t \\ \Longleftrightarrow & c_1(\xi) + b_1(\xi)(u^{(1)} - l^{(1)})\eta + a_1(\xi)(u^{(1)} - l^{(1)})^2\eta^2 \geq 0 \quad \forall\, 0 \leq \eta \leq 1, \xi \in \mathbf{R} \\ \Longleftrightarrow & \begin{cases} \begin{pmatrix} C_1 & (u^{(1)} - l^{(1)})B_1/2 - E_1 \\ (u^{(1)} - l^{(1)})B_1/2 - E_1^T & (u^{(1)} - l^{(1)})^2 A_1 + E_1 + E_1^T \end{pmatrix} \succeq 0, \\ E_1 + E_1^T \succeq 0, \\ \text{where } A_1 \in \mathfrak{A}, B_1 \in \mathfrak{B}(l^{(1)}, l^{(2)}, u^{(2)}), C_1 \in \mathfrak{C}(l^{(1)}, l^{(2)}, u^{(2)}). \end{cases} \end{aligned}$$

Hence, (D_2) is equivalent to the following SDP:

$$\begin{aligned} (SDP_2) \inf_{z_0, z_1, Z_2} \; & z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\ s.t. & \begin{cases} \begin{pmatrix} C_1 & (u^{(1)} - l^{(1)})B_1/2 - E_1 \\ (u^{(1)} - l^{(1)})B_1/2 - E_1^T & (u^{(1)} - l^{(1)})^2 A_1 + E_1 + E_1^T \end{pmatrix} \succeq 0, \\ E_1 + E_1^T \succeq 0 \\ A_1 \in \mathfrak{A}, B_1 \in \mathfrak{B}(l^{(1)}, l^{(2)}, u^{(2)}), C_1 \in \mathfrak{C}(l^{(1)}, l^{(2)}, u^{(2)}) \\ \begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} \succeq 0 \end{cases} \end{aligned}$$

7.2.3 Constraint (7.7) in LMIs

Let $x^{(1)} = u^{(1)} - \eta$ and $x^{(2)} = \frac{l^{(2)}+u^{(2)}\xi^2}{1+\xi^2}$. Writing the expression in (7.7) componentwise and multiplying $(1+\xi^2)^2$ on it, we have,

$$
\begin{aligned}
&(1+\xi^2)^2(z_0 - 1 + z_1^T x + Z_2 \bullet xx^T) \\
=&(1+\xi^2)^2(z_0 - 1 + z_1^{(1)}x^{(1)} + z_1^{(2)}x^{(2)} + Z_2^{(11)}(x^{(1)})^2 \\
&+ (Z_2^{(12)} + Z_2^{(21)})x^{(1)}x^{(2)} + Z_2^{(22)}(x^{(2)})^2) \\
=&\left(z_0 - 1 + z_1^{(1)}(u^{(1)} - \eta) + Z_2^{(11)}(u^{(1)} - \eta)^2\right)(1 + 2\xi^2 + \xi^4) \\
&+ \left[z_1^{(2)} + \left(Z_2^{(12)} + Z_2^{(21)}\right)(u^{(1)} - \eta)\right](l^{(2)} + u^{(2)}\xi^2)(1+\xi^2) \\
&+ Z_2^{(22)}(l^{(2)} + u^{(2)}\xi^2)^2 \\
=&c_2(\xi) + b_2(\xi)\eta + a_2(\xi)\eta^2,
\end{aligned}
$$

where

$$
\begin{aligned}
c_2(\xi) :=& \left(z_0 - 1 + z_1^{(1)}u^{(1)} + Z_2^{(11)}(u^{(1)})^2\right)(1+\xi^2)^2 \\
&+ \left[z_1^{(2)} + u^{(1)}\left(Z_2^{(12)} + Z_2^{(21)}\right)\right](l^{(2)} + u^{(2)}\xi^2)(1+\xi^2) \\
&+ Z_2^{(22)}(l^{(2)} + u^{(2)}\xi^2)^2 = \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}^T C_2 \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
C_2 \in & \mathfrak{C}(u^{(1)}, l^{(2)}, u^{(2)}) := \\
& \Big\{ Y \in \mathcal{S}^3 \mid Y^{(11)} = \Big(z_0 - 1 + z_1^{(1)} u^{(1)} + Z_2^{(11)} (u^{(1)})^2 \Big) \\
& + \Big[z_1^{(2)} + u^{(1)} \Big(Z_2^{(12)} + Z_2^{(21)} \Big) \Big] \, l^{(2)} + Z_2^{(22)} (l^{(2)})^2; \\
& Y^{(13)} + Y^{(22)} + Y^{(31)} = 2 \Big(z_0 - 1 + z_1^{(1)} u^{(1)} + Z_2^{(11)} (u^{(1)})^2 \Big) \\
& + \Big[z_1^{(2)} + u^{(1)} \Big(Z_2^{(12)} + Z_2^{(21)} \Big) \Big] \, (l^{(2)} + u^{(2)}) \\
& + 2 Z_2^{(22)} l^{(2)} u^{(2)}; \\
& Y^{(33)} = \Big(z_0 - 1 + z_1^{(1)} u^{(1)} + Z_2^{(11)} (u^{(1)})^2 \Big) \\
& + \Big[z_1^{(2)} + u^{(1)} \Big(Z_2^{(12)} + Z_2^{(21)} \Big) \Big] \, u^{(2)} + Z_2^{(22)} (u^{(2)})^2; \\
& Y^{(12)} = Y^{(21)} = Y^{(23)} = Y^{(32)} = 0 \Big\},
\end{aligned}
$$

$$
\begin{aligned}
b_2(\xi) := & - \Big(z_1^{(1)} + 2 Z_2^{(11)} l^{(1)} \Big) (1 + \xi^2)^2 \\
& - \Big(Z_2^{(12)} + Z_2^{(21)} \Big) (l^{(2)} + u^{(2)} \xi^2)(1 + \xi^2) \\
= & \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}^T B_2 \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
B_2 \in -\mathfrak{B}(u^{(1)}, l^{(2)}, u^{(2)}) :=& \\
&\Big\{ Y \in \mathcal{S}^3 \mid Y^{(11)} = -\left(z_1^{(1)} + 2Z_2^{(11)} u^{(1)}\right) - \left(Z_2^{(12)} + Z_2^{(21)}\right) l^{(2)}; \\
&Y^{(13)} + Y^{(22)} + Y^{(31)} = -2\left(z_1^{(1)} + 2Z_2^{(11)} u^{(1)}\right) \\
&- \left(Z_2^{12} + Z_2^{21}\right)(l^{(2)} + u^{(2)}); \\
&Y^{(33)} = -\left(z_1^{(1)} + 2Z_2^{(11)} u^{(1)}\right) + \left(Z_2^{(12)} + Z_2^{(21)}\right) u^{(2)}; \\
&Y^{(12)} = Y^{(21)} = Y^{(23)} = Y^{(32)} = 0 \Big\},
\end{aligned}
$$

$$
a_2(\xi) := Z_2^{(11)}(1+\xi^2)^2 = \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}^T A_2 \begin{pmatrix} 1 \\ \xi \\ xi^2 \end{pmatrix},
$$

$$
\begin{aligned}
A_2 \in \mathfrak{A} :=& \{ Y \in \mathcal{S}^3 \mid Y^{(11)} = Z_2^{(11)}; Y^{(13)} + Y^{(22)} + Y^{(31)} = 2Z_2^{(11)}; \\
& Y^{(33)} = Z_2^{(11)}; Y^{(12)} = Y^{(21)} = Y^{(23)} = Y^{(32)} = 0 \}.
\end{aligned}
$$

By Theorem 13, constraint (7.7) can be represented by LMIs:

$$
\begin{aligned}
& z_0 + z_1^T x + Z_2 \bullet x x^T \geq 0 \quad \forall -\infty < x^{(1)} < u^{(1)}, -\infty < x^{(2)} < u^{(2)} \\
\Longleftrightarrow & c_2(\xi) + b_2(\xi)\eta + a_2(\xi)\eta^2 \geq 0 \quad \forall\, \eta \geq 0, \xi \in \mathbf{R} \\
\Longleftrightarrow & \left\{ \begin{array}{l} \begin{pmatrix} C_2 & B_2/2 \\ B_2/2 & A_2 \end{pmatrix} - \begin{pmatrix} 0 & E_2 \\ E_2^T & 0 \end{pmatrix} \succeq 0, E_2 + E_2^T \succeq 0 \\ \text{where } A_2 \in \mathfrak{A}, B_2 \in -\mathfrak{B}_2(u^{(1)}, l^{(2)}, u^{(2)}), C_2 \in \mathfrak{C}(u^{(1)}, l^{(2)}, u^{(2)}). \end{array} \right.
\end{aligned}
$$

Summarizing the results, the tractable formulation for the

(D_3) is

$$
(SDP_3) \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T)
$$

$$
s.t. \begin{cases} \begin{pmatrix} C_2 & B_2/2 \\ B_2/2 & A_2 \end{pmatrix} - \begin{pmatrix} 0 & E_2 \\ E_2^T & 0 \end{pmatrix} \succeq 0, \\ E_2 + E_2^T \succeq 0 \\ A_2 \in \mathfrak{A}, B_2 \in -\mathfrak{B}(u^{(1)}, l^{(2)}, u^{(2)}), C_2 \in \mathfrak{C}(u^{(1)}, l^{(2)}, u^{(2)}), \\ \begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} \succeq 0 \end{cases}
$$

7.3 Several model extensions

7.3.1 Moment bound of probability of union events

For the tight bound of the union of two events, say $\sup_{x\sim(\mu,\sigma)} \mathbb{P}(l^{(1)} \leq x^{(1)} \leq u^{(1)}$ or $x^{(2)} \leq u^{(2)})$, applying Theorems 13 and 14 to its dual is almost immediate:

$$
\inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T)
$$

$$
s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall l^{(1)} \leq x^{(1)} \leq u^{(1)} \tag{7.8}
$$

$$
z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall x^{(2)} \leq u^{(2)} \tag{7.9}
$$

$$
z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x \in \mathbf{R}^2 \tag{7.10}
$$

We can see that Theorem 14 is applied to (7.8) and Theorem 13 to (7.9).

7.3.2 The variety of domain of x

We can extend to compute the corresponding bound of (P_k), $k = 1, 2, 3$, for nonnegative random variables. The dual formulations are respectively

$$
\begin{aligned}
(D_{1+}) \quad & \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall 0 \leq x \leq u \qquad (7.11) \\
& \qquad\;\; z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x^{(1)}, x^{(2)} \in \mathbf{R}_+ \\
(D_{2+}) \quad & \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \quad \forall \max\{0, l\} \leq x \leq u \qquad (7.12) \\
& \qquad\;\; z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x^{(1)}, x^{(2)} \in \mathbf{R}_+ \\
(D_{3+}) \quad & \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \quad z_0 + z_1^T x + Z_2 \bullet xx^T \geq 1 \\
& \qquad\qquad \forall 0 \leq x^{(1)} \leq u^{(1)}, \max\{0, l^{(2)}\} \leq x^{(2)} \leq u^{(2)} \qquad (7.13) \\
& \qquad\;\; z_0 + z_1^T x + Z_2 \bullet xx^T \geq 0 \quad \forall x^{(1)}, x^{(2)} \in \mathbf{R}_+
\end{aligned}
$$

Approximations to (D_{1+}) with sos polynomials are discussed in Cox et al. [12]. Theorems 14 can be applied to (7.11), (7.12) and (7.13) in the same way as that in the previous section. Their second constraint are now a copositive constraint of dimension 3×3, i.e.,

$$\begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} \in \mathcal{C}^3,$$

and therefore an LMI. Then the SDP for (D_{k+}), $k = 1, 2, 3$, are

$$(SDP_{1+}) \qquad \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T)$$

$$s.t. \begin{cases} \begin{pmatrix} z_0 - 1 + z_1^T u + u^T Z_2 u & -z_1^T/2 - u^T Z_2 \\ -z_1/2 - Z_2 u & Z_2 \end{pmatrix} - N_1 \in \mathcal{S}_+^3, \\ N_1^{(ij)} \geq 0, \quad i, j = 1, 2, 3 \\ \begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} - N_2 \in \mathcal{S}_+^3 \\ N_2^{(ij)} \geq 0, \quad i, j = 1, 2, 3 \end{cases}$$

$$
(SDP_{2+})\ \inf_{z_0,z_1,Z_2} z_0 + z_1^T\mu + Z_2 \bullet (\Gamma + \mu\mu^T)
$$

$$
s.t.\begin{cases}
\begin{pmatrix} C_1 & b^*B_1/2 - E_1 \\ b^*B_1/2 - E_1^T & a^*A_1 + E_1 + E_1^T \end{pmatrix} \succeq 0, \\
\text{where } b^* = (u^{(1)} - \max\{l^{(1)},0\}), a^* = (u^{(1)} - \max\{l^{(1)},0\})^2 \\
E_1 + E_1^T \succeq 0 \\
A_1 \in \mathfrak{A}, B_1 \in \mathfrak{B}(\max\{l^{(1)},0\}, \max\{l^{(2)},0\}, u^{(2)}), \\
C_1 \in \mathfrak{C}(\max\{l^{(1)},0\}, \max\{l^{(2)},0\}, u^{(2)}) \\
\begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} - N \in \mathcal{S}_+^3 \\
N^{(ij)} \geq 0, \quad i,j = 1,2,3
\end{cases}
$$

$$
(SDP_{3+})\qquad \inf_{z_0,z_1,Z_2} z_0 + z_1^T\mu + Z_2 \bullet (\Gamma + \mu\mu^T)
$$

$$
s.t.\begin{cases}
\begin{pmatrix} C_1 & u^{(1)}B_1/2 - E_1 \\ u^{(1)}B_1/2 - E_1^T & (u^{(1)})^2A_1 + E_1 + E_1^T \end{pmatrix} \succeq 0, \\
E_1 + E_1^T \succeq 0 \\
A_1 \in \mathfrak{A}, B_1 \in \mathfrak{B}(0, \max\{l^{(2)},0\}, u^{(2)}), \\
C_1 \in \mathfrak{C}(0, \max\{l^{(2)},0\}, u^{(2)}) \\
\begin{pmatrix} z_0 & z_1^T/2 \\ z_1/2 & Z_2 \end{pmatrix} - N \in \mathcal{S}_+^3 \\
N^{(ij)} \geq 0, \quad i,j = 1,2,3
\end{cases}
$$

Meanwhile, it is worth noting that the results above imply that we can compute the joint probability bound for two random variables with a variety of support of $x^{(1)}$ and $x^{(2)}$, in any combination:

(I) $x \in \mathbf{R}$;

(II) $x \in \mathbf{R}_+$ (therefore any semi-infinite interval will do.);

(III) $x \in [a, b]$ for any constant a and b;

(IV) $x \in \bigcup_{j=1}^{\bar{j}} I_j$, where I_j's can be intervals of above forms.

7.3.3 Higher moments incorporated

We have shown that Theorems 13 and 14 are adapted to the computation of joint probability bound perfectly. As a matter of fact, the theorems provide us with the freedom of using either random variable's higher moments. For example, for the bound of $\mathbb{P}(l^{(1)} \leq x^{(1)} \leq u^{(1)}, l^{(2)} \leq x^{(2)} \leq u^{(2)})$, if we are given the higher moments of $x^{(2)}$, say $\lambda_3, \cdots, \lambda_n$, in addition to μ and Γ, then the dual formulation is

$$\inf_{z_0, z_1, Z_2, z_3, \cdots, z_n} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) + \sum_{i=3}^{n} z_i \lambda_i$$

$$s.t. \begin{cases} z_0 + z_1^T x + Z_2 \bullet xx^T + \sum_{i=3}^{n} z_i (x^{(2)})^i \geq 1 & \forall l \leq x \leq u \\ z_0 + z_1^T x + Z_2 \bullet xx^T + \sum_{i=3}^{n} z_i (x^{(2)})^i \geq 0 & \forall x^{(1)}, x^{(2)} \in \mathbf{R} \end{cases}$$

Every step in the previous setting can be applied and the dimension m in Theorems 13 and 14 is now chosen as $m = n+1$.

7.4 Applications of the moment bound

7.4.1 The Riemann integrable set approximation

Given any bounded Riemann integrable subset $\mathcal{R}$ of a sample space $\Omega \subset \mathbf{R}^2$ and the first two moments (μ and Γ) of an arbitrary probability measure of it, we can always approximate the distributional robust probability measure of $\mathcal{R}$ with finitely many rectangular partitions $[\underline{x}_i, \overline{x}_i] \times [\underline{y}_i, \overline{y}_i]$. In other words, there exists $m \in \mathbf{N}$ such that $\mathcal{R} \approx \bigcup_{i=1}^{m} [\underline{x}_i, \overline{x}_i] \times [\underline{y}_i, \overline{y}_i]$ and

$$
\begin{aligned}
& \sup_{(x,y)\sim(\mu,\sigma)} \mathbb{P}\left((x,y) \in \mathcal{R}\right) \\
\approx & \sup_{(x,y)\sim(\mu,\sigma)} \mathbb{P}\left((x,y) \in \bigcup_{i=1}^{m} [\underline{x}_i, \overline{x}_i] \times [\underline{y}_i, \overline{y}_i]\right) \\
= & \inf_{z_0, z_1, Z_2} z_0 + z_1^T \mu + Z_2 \bullet (\Gamma + \mu\mu^T) \\
& s.t. \begin{cases} z_0 + z_1^T x + Z_2 \bullet x x^T \geq 1 & \forall x \in [\underline{x}_i, \overline{x}_i] \times [\underline{y}_i, \overline{y}_i], i = 1, \cdots, m \\ z_0 + z_1^T x + Z_2 \bullet x x^T \geq 0 & \forall x, y \in \mathbf{R} \end{cases}
\end{aligned}
$$

7.4.2 Worst-case simultaneous VaR

VaR refers to the risk of a single asset or a whole portfolio. Given the international investment markets nowadays, dependence among various seemingly unrelated factors have given rise to a growing concern. Therefore, it is of great importance to study the VaR of different portfolios simultaneously. Consider two investment markets ($i = 1, 2$). Suppose that $\theta_i \in \mathbf{R}^{p_i}$ be the vector of investment return from p_i assets with a mean $m_i \in \mathbf{R}^{p_i}$, second moment matrix $M_i \in \mathcal{S}_+^{p_i}$, and covariance matrix between the two markets $C_{12} \in \mathbf{R}^{p_1 \times p_2}$. Let $w_i \in \mathbf{R}^{p_i}$ be the portfolio weights and $x^{(i)} = -w_i^T\theta_i$ the portfolio return. If $\alpha^{(i)}$ is the VaR of the portfolio $w_i^T\theta_i$ in the two markets, then we can compute the worst-case probability through (SDP_1), in which we take $\mu = [-w_1^T m_1, -w_2^T m_2]^T$, $\Gamma = \begin{pmatrix} w_1^T M_1 w_1 & w_1^T C_{12} w_2 \\ w_1^T C_{12} w_2 & w_2^T M_2 w_2 \end{pmatrix}$ and $u = [\alpha^{(1)}, \alpha^{(2)}]^T$. In fact, (SDP_1) allows us to compute the probability exactly (upon machine error) for at most three portfolios under (SDP_1).

In general, when we let $\alpha = \alpha^{(1)} = \ldots = \alpha^{(d)}$, we can define the *worst-case simultaneous VaR (WS-VaR)* by

$$\begin{aligned} &WS\text{-}VaR_\epsilon(w_1^T\theta_1, \ldots, w_d^T\theta_d) \\ &:= \arg\min_\alpha \{\mathbb{F}_{d,2}(-w_1^T\theta_1 \geq \alpha, \ldots, -w_d^T\theta_d \geq \alpha) \leq \epsilon\}, \end{aligned}$$

Since $\mathbb{F}_{d,2}(-w_1^T\theta_1 \geq \alpha, \ldots, -w_d^T\theta_d \geq \alpha)$ is monotone in α, WS-VAR can be obtained by line search methods.

7.5 Conclusion

We have introduced the concept of distributional robust probability function for moment bound probability to generalize the framework. In particular, $\mathbb{F}_{1,2}(x \in \mathbf{R} : x \leq t)$ turns out to be a probability itself while this may not be so in general. This result applied in portfolio selection with VaR minimization also matches that of El Ghaoui et al. [21]. When the two dimensional random variable is considered, we have introduced the computation methodology, which is mainly due to Luo et al. [50], for a rather comprehensive collection of events. Under mild assumptions, we also propose the idea of worst-case simultaneous VaR and take into account the risk of two or three portfolios at the same time. Such quantities clearly have an important role to play in risk management.

□ **End of chapter.**

Chapter 8

Concluding Remarks and Future Directions

There are several potential future work following the results in this thesis. In view of the theoretical development of the S-Lemma, three interesting directions can be explored. One is to study the corresponding version of the S-Lemma, if it exists, in the setting of bivariate quartic polynomials. This motivation comes from the parallel comparison between the existence of sos-polynomials-certificate for nonnegative polynomials (according to Hilbert in 1888) and that of the S-Lemma. Another similar pursuit is to study the existence of S-Lemma in complex polynomials. This may not even be a well-posed statement, but it is definitely of a general research interest. Last but not least, we are keen on verifying if the S-Lemma in the univariate polynomials can accommodate more functions $g(x)$'s. In other words, we study the equivalence between Statement 1 and Statement 3 for $k \geq 2$. This equivalence will be surprising on the ground that the original S-Lemma only holds for a single function $g(x)$. On the other hand, continuing efforts will be made for discovering applications wherever appropriate.

□ **End of chapter.**

Appendix A

Nonnegative univariate polynomials

Theorem 15. *(cf. Theorem 17.10 in Nesterov [56]) Let* $p(x) = \sum_{i=1}^{2d} p_i x^i$ *be a univariate polynomial of degree* $2d$*.* $p(x) \geq 0$ *for all* $x \in \mathbf{R}$ *if and only if there exists a positive semidefinite matrix* Q *(in symbol we write* $Q \succeq 0$*) such that*

$$p_i = \sum_{j+k=i} Q_{jk},$$

where Q_{jk} *is the entry of* Q *in the* j*'s row and* k*'s column.*

Theorem 16. *(cf. Theorem 17.11 in Nesterov [56]) Let* $p(x) = \sum_{i=1}^{d} p_i x^i$ *be a univariate polynomial of degree* d*.* $p(x) \geq 0$ *for all* $x \in \mathbf{R}_+$ *can be represented by an LMI.*

Proof. Take $x = y^2$ and apply Theorem 15. $\square$

Theorem 17. *(cf. Theorem 17.12 in Nesterov [56]) Let* $p(x) = \sum_{i=1}^{d} p_i x^i$ *be a univariate polynomial of degree* d*.* $p(x) \geq 0$ *for all* $x \in [a, b]$ *can be represented by an LMI.*

Proof. Take $x = (b-a)\frac{y^2}{y^2+1} + a$ and apply Theorem 15 on $(y^2 + 1)p((b-a)\frac{y^2}{y^2+1} + a) \geq 0$. $\square$

$\square$ **End of chapter.**

Appendix B

First and second moment of (7.2)

We can regard $F(t)$ in 7.2 as a distribution of t with density function

$$\frac{d}{dt}F(t) = f(t) = \begin{cases} \frac{2\sigma^2(\mu-t)}{[(\mu-t)^2+\sigma^2]^2}, & \mu > t \\ 0, & \mu \leq t. \end{cases}$$

Then

$$\begin{aligned}
\mathbb{E}(t) =& \int_{-\infty}^{\mu} \frac{2\sigma^2(\mu-t)t}{[(\mu-t)^2+\sigma^2]^2}dt \\
=& 2\sigma^2 \left[\int_{-\infty}^{\mu} \frac{-(\mu-t)^2}{[(\mu-t)^2+\sigma^2]^2}dt + \int_{-\infty}^{\mu} \frac{\mu(\mu-t)}{[(\mu-t)^2+\sigma^2]^2}dt\right] \\
=& 2\sigma^2 \left[\int_{\pi/2}^{0} \frac{\sigma^3 \tan^2\theta \sec^2\theta}{\sigma^4 \sec^4}d\theta + \int_{\pi/2}^{0} \frac{-\sigma^2\mu \tan\theta \sec^2\theta d\theta}{\sigma^4 \sec^4\theta}d\theta\right] \\
=& 2\sigma^2 \left[\frac{1}{\sigma}\int_{\pi/2}^{0} \sin^2\theta d\theta - \frac{\mu}{\sigma^2}\int_{\pi/2}^{0} \sin\theta\cos\theta d\theta\right] \\
=& 2\sigma^2 \left[\frac{1}{2\sigma}\int_{\pi/2}^{0} \frac{1-\cos 2\theta}{2}d(2\theta) - \frac{\mu}{4\sigma^2}\int_{\pi/2}^{0} \sin 2\theta d(2\theta)\right] \\
=& 2\sigma^2 \left[\frac{1}{2\sigma}\left[\theta - \frac{\sin 2\theta}{2}\right]_{\pi/2}^{0} - \frac{\mu}{4\sigma^2}\left[-\cos 2\theta\right]_{\pi/2}^{0}\right] \\
=& \mu - \frac{\pi}{2}\sigma
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(t^2) =& \int_{-\infty}^{\mu} \frac{2\sigma^2(\mu-t)t^2}{[(\mu-t)^2+\sigma^2]^2}dt \\
=& 2\sigma^2 \int_{0}^{\pi/2} \frac{\sigma\tan\theta(\mu-\sigma\tan\theta)^2}{\sigma^4\sec^4\theta}\sigma\sec^2\theta d\theta \\
=& 2\int_{0}^{\pi/2} \tan\theta(\mu\cos\theta - \sigma\sin\theta)^2 d\theta \\
=& 2\int_{0}^{\pi/2} (\mu^2\sin\theta\cos\theta + \sigma^2\tan\theta\sin^2\theta - 2\mu\sigma\sin^2\theta)d\theta \\
=& 2\int_{0}^{\pi/2} ((\mu^2-\sigma^2)\sin\theta\cos\theta - 2\mu\sigma\sin^2\theta)d\theta + 2\sigma^2\int_{0}^{\pi/2} \tan\theta d\theta
\end{aligned}$$

We can check that the integral is finite while the second is infinite. Hence, $\mathbb{E}(t^2) = +\infty$

□ **End of chapter.**

Bibliography

[1] Ai, W. , Huang, Y. , and Zhang, S. . New results on hermitian matrix rank-one decomposition. *Mathematical Programming, Series A*, 128:253–283, 2011.

[2] Ben-Tal, A. , El Ghaoui, L. , and Nemirovski, A. . *Robust Optimization.* Princeton University Press, 2009.

[3] Ben-Tal, A. , Goryashki, A. , Guslitzer, E. , and Nemirovski, A. . Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99:351–376, 2004.

[4] Ben-Tal, A. and Nemirovski, A. . *Lectures on Modern Convex Optimization: Analysis, Algoritms, and Engineering Applications, MPS/SIAM Series on Optimization 2.* SIAM, Philadephia, 2001.

[5] Ben-Tal, A. and Nemirovski, A. . On tractable approximations of uncertain linear matrix inequalities affected by in-

terval uncertainty. *SIAM Journal on Optimization*, 12:811–833, 2002.

[6] Bertsimas, D. and Popescu, I. . Optimal inequalities in probability thoery: A convex optimization approach. *SIAM Journal of Optimization*, 15(3):780–804, 2005.

[7] Boyd, S. and Vandenberghe, L. . *Convex optimization.* Cambridge University Press, 2004.

[8] Brockett, P. L. and Cox, S. H. . Insurance calculations using incomplete information. *Scandinavian Actuarial Journal*, pages 94–108, 1985.

[9] Chebyshev, P. . Sur les valeurs limites des integrales. *J. Mah Pures Appl*, 19(2):157–60, 1874.

[10] Chen, L. , He, S. , and Zhang, S. . Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, 59(4):847–865, 2011.

[11] Cox, S. H. . Bounds on expected values of insurance payments and option prices. *Transactions of the Society of Actuaries*, 43:231–260, 1991.

[12] Cox, S. H. , Lin, Y. , Tian, R. , and Zuluaga, L. F. . Bounds for ruin probabilities and value at

risk. *Actuarial Research Clearing House Newsletters 1*, 2008. `http://library.soa.org/library/proceedings/arch/2008/arch-2008-iss1-cox-lin-tian.pdf`.

[13] De Schepper, A. and Heijnen, B. . Distribution-free option pricing. *Insurance: Mathematics and Economics*, 40(2):179–199, 2007.

[14] De Schepper, A. and Heijnen, B. . How to estimate the value at risk under incomplete information. *Journal of Computational and Applied Mathematics*, 233(9):2213–2226, 2010.

[15] De Vylder, F. . Best upper bounds for integrals with respect to measures allowed to vary under conical and integral constraints. *Insurance: Mathematics and Economics*, 1(2):109–130, 1982.

[16] De Vylder, F. . Bound on intergrals: Eliminiation of the dual and reduction of the number of equality constraints. *Insurance: Mathematics and Economics*, 2(3):139–145, 1983.

[17] De Vylder, F. . Maximization, under equality constraints, of a functional of a probability distribution. *Insurance: Mathematics and Economics*, 2(1):1–16, 1983.

[18] De Vylder, F. and Goovaerts, M. . Analytical best upper bounds on stop-loss premiums. *Insurance: Mathematics and Economics*, 1(3):197–211, 1982.

[19] De Vylder, F. and Goovaerts, M. . Maximization of the variance of a stop-loss reinsured risk. *Insurance: Mathematics and Economics*, 2(2):75–80, 1983.

[20] Dines, L. L. . On the mapping of quadratic forms. *Bulletin of the American Mathematical Society*, 47(6):=494–498, 1941.

[21] El Ghaoui, L. , Oks, M. , and Oustry, F. . Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[22] Frey, J. . Optimal distribution-free confidence bands for a distribution function. *Journal of Statistical Planning and Inference*, 138, 2009.

[23] Goovaerts, M. , Haezendonck, J. , and De Vylder, F. . Numerical best bounds on stop-loss premiums. *Insurance: Mathematics and Economics*, 1(4):287–302, 1982.

[24] Goovaerts, M. and Kaas, R. . Application of the problem of moments to derive bounds on integrals with inte-

gral constraints. *Insurance: Mathematics and Economics*, 4(2):99–111, 1985.

[25] Goovaerts, M. J. , Kaas, R. , and Laeven, R. J. A. . Worst case risk measurement: Back to the future? *Insurance: Mathematics and Economics*, 49(3):380–392, 2011.

[26] Grant, M. and Boyd, S. . CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx/`, apr 2011.

[27] Grundy, B. . Option prices and the underlying asset's return distribution. *Journal of Finance*, 46(3):1045–1069, 1991.

[28] Han, D. , Li, X. , Sun, D. , and Sun, J. . Bounding option prices of multi-assets: A semidefinite programming approach. *Pacific Journal of Optimization*, 1(1):59–79, 2005.

[29] He, S. , Zhang, J. , and Zhang, S. . Bounding probability of small deviation: a fourth moment approach. *Mathematics of Operationa Research*, 35(1):208–232, 2010.

[30] Hoel, P. G. . Confidence regions for linear regression. *Proceedings of the second Berkeley Symposium*, 1951.

[31] Hu, S.-L. and Huang, Z.-H. . Theorems of the alternative for inequality systems of real polynomials. *Jour-*

nal of Optimization Theory and Applications, 2012. DOI 10.1007/s10957-012-9993-9.

[32] Isii, K. . On sharpness of tchebycheff-type inequalities. *Annals of the institute of statistical mathematics*, 14(1):185–197, 1962.

[33] Jansen, K. , Haezendonck, J. , and Goovaerts, M. . Upper bounds on stop-loss premiums in case of known moments up to the fourth order. *Insurance: Mathematics and Economics*, 5(4):315–334, 1986.

[34] Jonsson, U. T. . A lecture on the s-procedure, 2001.

[35] Kaas, R. and Goovaerts, M. J. . Best bounds for positive distributions with fixed moments. *Insurance: Mathematics and Economics*, 5(1):87–92, 1986.

[36] Karlin, S. and Studden, W. . *Tchebycheff systems: With applications in analysis and statistics.* Pure and Applied Mathematics, a Series of Texts and Monographs. Interscience Publishers, John Wiley and Sons, New York, 1966.

[37] Kemperman, J. H. B. . The general moment problem, a geometric approach. *Annals of Mathematical Statistics*, 39(1):93–122, 1968.

[38] Kleywegt, A. , Shapiro, A. , and Homen-De-Mello, T. . The sample average approximation methods for stochastic discrete optimization. *SIAM Journal of Optimization*, 12(2):479–502, 2001.

[39] Lasserre, J.-B. . *Moments, Positive Polynomials, and Their Applications, Imperial College Press Optimization Series Vol. 1.* Imperial College Press, London, 2010.

[40] Lasserre, J. B. , Preito-Rumeau, T. , and Zervos, M. . Pricing a class of exotic options via moments and sdp relaxations. *Mathematical Finance*, 16(3):469–494, 2006.

[41] Laurence, P. and Wang, T.-H. . Sharp distribution free lower bounds for spread options and the corresponding optimal subreplicating portfolios. *Insurance: Mathematics and Economics*, 41(1):35–47, 2009.

[42] Levy, H. . Upper and lower bounds of put and call option value: Stochastic dominance approach. *Journal of Finance*, 40:1197–1217, 1985.

[43] Linderoth, J. , Shapiro, A. , and Wright, S. . The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142(1):215–241, 2006.

[44] Liu, G. and Li, W. V. . Moment bounds for truncated random variables. *Statistics and Probability Letters*, 79:1951–1956, 2009.

[45] Liu, W. . *Simultaneous inference in regression.* CRC Press, Taylor & Group, 2011.

[46] Liu, W. , Lin, S. , and Piegorsch, W. . Construction of exact simultaneous confidence bands for a simple linear regression model. *International statistical review*, 76, 2008.

[47] Liu, W. , Wynn, H. , and Hayter, A. . Statistical inferences for linear regression models when the covariates have functional relationships: polynomial regression. *Journal of statistical computation and simulation*, 78, 2008.

[48] Lo, A. . Semiparametric upper bounds for option prices and expected payoff. *Journal of Financial Economics*, 19:373–388, 1987.

[49] Luo, Z. Q. , Sturm, J. F. , and Zhang, S. . Duality and self-duality for conic convex programming. Technical Report 9620/A, Econometric Institute, Erasmus University, Rotterdam, 1996.

[50] Luo, Z. Q. , Sturm, J. F. , and Zhang, S. . Multivariate nonnegative quadratic mappings. *SIAM Journal on optimization*, 14(4):1140–1162, 2004.

[51] Markov, A. . *On certain applications of algebraic continued fractions*. PhD thesis, St. Petersburg, 1884. (in Russian).

[52] Naiman, D. . Conservative confidence regions in curvilinear regression. *Annals of mathematical statistics*, 14, 1986.

[53] Natarajan, K. and Sim, M. . Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, 20(4):695–731, 2010.

[54] Nemirovski, A. . Lectures on modern convex optimization. `http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf`.

[55] Nemirovski, A. . Lectures on robust convex optimization. `http://www2.isye.gatech.edu/~nemirovs`.

[56] Nesterov, Y. . Squared functional systems and optimization problems. In Frenk, H. , Roos, K. , Terlaky, T. , and Zhang, S. , editors, *High Performance Optimization Techniques*, chapter 17. Kluwer Academic Publishers, 2000.

[57] Pólik, I. and Terlaky, T. . A survey of the s-lemma. *SIAM Review*, 49(3):371–418, 2007.

[58] Polya, G. and Szego, G. . *Problems and Theorems in Analysis II.* Springer-Verlag, 1976.

[59] Polyak, B. T. . Convexity of quadratic transformations and its use in control and optimization. *Journal of optimization theory and applications*, 99(3):553–583, 1998.

[60] Popescu, I. . A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657, 2005.

[61] Popescu, I. . Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[62] Rodriguez, R. J. . Option pricing bounds: synthesis and extension. *Journal of Financial Research*, XXVI(2):149–164, 2003.

[63] Scarf, H. . A min-max solution of an inventory problem. In Arrow, K. , Karlin, S. , and Scarf, H. , editors, *Studies in the Mathematical Theory of Inventory and Production.* Stanford University Press, 1958.

[64] Scheffe, H. . A method for judging all contrasts in analysis of variance. *Biometrika*, 40, 1953.

[65] Scheffe, H. . *The analysis of variance*. Wiley, 1959.

[66] Shapiro, A. . On duality theory of conice linear problems. In Goberna, M. A. and Lópes, M. A. , editors, *Semi-infinite Programming - recent advances*, chapter 7. Kluwer Academic Publishers, 2001.

[67] Shapiro, A. and Homen-De-Mello, T. . On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. *SIAM Journal of Optimization*, 11(1):70–86, 2000.

[68] Spurrier, J. . Comparison of simultaneous confidence bands for quadratic regression over a finite interval. *Technometrics*, 35, 1993.

[69] Sturm, J. F. . Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones (updated for version 1.05). `http://sedumi.ie.lehigh.edu/index.php?option=com_docman&task=doc_download&gid=5&Itemid=76`, 2004.

[70] Sturm, J. F. and Zhang, S. . On cones of nonnegative quadratic functions. *Mathematics of Operations Research*, 28:246–267, 2003.

[71] Taylor, G. C. . Upper bounds on sstop-loss premiums under constraints on claim size distribution. *Scandinavian Actuarial Journal*, pages 94–105, 1977.

[72] Taylor, G. C. . Discussion of best 'bounds for positive distributions with fixed moments' by r. kaas and m.j. goovaerts. *Insurance: Mathematics and Economics*, 5(1):93–95, 1986.

[73] Tian, R. . *Moment problems with applications to Value-at-Risk and portfolio management.* PhD thesis, Georgia State University, 7 2008.

[74] Tuy, H. and Tuan, H. D. . Generalized s-lemma and strong duality in nonconvex quadratic programming. `http://www.math.ac.vn/library/download/e-print/11/pdf/htuy1101.pdf`, 2011.

[75] UUsipaikka, E. . Exact confidence bands for linear-regression over intervals. *Journal of the American Statistical Association*, 78, 1983.

[76] Wong, M. H. and Zhang, S. . Computing best bounds for nonlinear risk measures with partial information.

[77] Working, H. and Hotelling, H. . Applications of the theory of error to the interpretation of trends. *Journal of the American Statistical Association*, 24, 1929.

[78] Wynn, H. . An exact confidence band for one-dimensional polynomial regression. *Biometrika*, 71(2), 1984.

[79] Wynn, H. and Bloomfield, P. . Simultaneous confidence bands in regression analysis. *Journal of the royal statistical society (B)*, 71(2), 1984.

[80] Yakubovich, A. V. . S-procedure in nonlinear control theory. *Vestnik Leningrad University*, 1:62–77, 1971.

[81] Yakubovich, V. A. . S-procedure in nonlinear control theory. *Vestnik Leningrad University*, 4:62–77, 1977.

[82] Yuan, Y. . On a subproblem of trust region algorithms for constrained optimization. *Mathematical Programming*, 47:53–63, 1990.

[83] Zuluaga, L. and Peña, J. . A conic programming approach to generalized tchebycheff inequalities. *Mathematics of Operations Research*, 30(2):369–388, 2005.

[84] Zuluaga, L. F. , Peña, J. , and Du, D. . Third-order extensions of lo's semiparametric bound for european call options. *European Journal of Operational Research*, 198:557–570, 2009.